University of Namur

THESIS / THÈSE

MASTER IN COMPUTER SCIENCE

Website navigability automated assessment
empirical validation of a multi-level quality model

Moulart, Antoine

Award date:
2011

Awarding institution:
University of Namur

Facultés Universitaires Notre-Dame de la Paix
Faculté d'Informatique
Année académique 2010-2011

Website Navigability Automated Assessment

Empirical Validation of a Multi-level Quality Model

Antoine Moulart

Mémoire présenté en vue de l'optention du grade de Master en Informatique.

Facultés Universitaires Notre-Dame de la Paix
Faculté d'Informatique
Année académique 2010-2011

Website Navigability Automated Assessment

Empirical Validation of a Multi-level Quality Model

Antoine Moulart

Promoteur du mémoire :
Naji Habra
FUNDP

Maître de stage :
Houari Sahraoui
Université de Montréal

Mémoire présenté en vue de l'optention du grade de Master en Informatique.

Abstract

English summary

The Web is an essential way to search for information. In the professional world, a website reflects the corporate image and is an important and efficient means of communication in order to interact with customers. The most effective websites are those with a high level of usability (i.e. the effectiveness with which a user can achieve his goals). Therefore, research has studied ways to assess how websites should be structured to ensure their quality and usability. We review the scientific literature about the assessment of website quality and usability. We focus on an important characteristic of the usability: the website navigability (i.e. the ease with which users can locate and access relevant information). Most research has focused on assessing the quality of individual pages or of a site as a whole.
Based on two previous studies, we propose the use of a multi-level quality model, applied to the problem of assessing the website navigability. Our Multi-level Model computes a navigability score for each webpage (page-level model), an aggregated score for all the pages (composition model), and the website navigability score (site-level model). Our Multi-level Model combines these three "cascading models". We contribute to the validation and comparison of these models.
To test our models, we conducted an experiment in two phases, with 22 and then 24 subjects. We found that the multi-level model is better predictor of navigability. We investigated several possible composition models. We tested different weighting strategies based on known graph analysis algorithms. We found that a complex weighting strategy does not improve the performance of our models. Finally, we propose different ways to improve our models, according to the results of the experiment.
Keywords : Website navigability, Website improvements, Quality assessment, Empirical study, Quality model

Résumé en français

Le Web est aujourd'hui un moyen incontournable de rechercher de l'information. Dans le monde professionnel, le site Web reflète l'image de l'entreprise et constitue un moyen de communication efficace pour interagir avec les clients. Les sites Web les plus efficaces sont ceux qui présentent un bon niveau d'utilisabilité (l'efficacité avec laquelle un utilisateur peut atteindre son but). Par conséquent, le monde de la recherche a étudié différents moyens d'évaluer la structure d'un site Web, afin d'assurer sa qualité et son utilisabilité. Nous passons en revue la littérature scientifique concernant l'évaluation de la qualité Web, et plus précisément de l'utilisabilité. Nous étudions en particulier une sous-caractéristique importante de l'utilisabilité : la navigabilité (la facilité avec laquelle un utilisateur peut localiser et accéder à l'information pertinente). La majorité des recherches s'est focalisée sur l'évaluation de la qualité des pages prises individuellement ou d'un site pris en entier.
Sur base de deux précédentes recherches, nous proposons ici l'utilisation d'un modèle qualité multi-niveaux, appliqué au problème de la navigabilité d'un site Web. Notre "modèle multi-niveaux" calcule un score pour chaque page web (modèle niveau-page), un score agrégé pour la qualité des pages (modèle de composition) et le score du site web (modèle niveau-site). Notre "modèle multi-niveaux" combine ces trois modèles, évalués "en cascade". Nous contribuons à la validation et à la comparaison de ces modèles. Pour tester nos modèles, nous menons une expérimentation en deux phases avec 22 puis 24 participants. Nous avons découvert que le "modèle multi-niveaux" est un meilleur indicateur de navigabilité. Nous avons examiné différents modèles de composition (en testant différents algorithmes) permettant de produire une valeur agrégée de la qualité des pages. Il apparait qu'une méthode de pondération complexe n'améliore pas la performance du modèle. Suite aux résultats de l'expérimentation, nous proposons différentes pistes d'amélioration.
Mots clés: Navigabilité des sites web, évaluation de la qualité, étude empirique, modèle qualité

Acknowledgments

First, I greatly thank my supervisor, Dr. Naji Habra, for his support and advice during this academic year. I thank you for allowing me to study an interesting topic and have a rewarding internship in Montréal. Thank you for your comments, critiques, and feedback about my Master's Thesis.

I would like to thank Dr. Houari Sahraoui for welcoming me at the University of Montréal. I am very grateful for your help and support during my stay in Montréal.
I would also like to thank Dr. Stéphane Vaucher for his explanations and suggestions during my research work. Thank you for taking the time to review a chapter.

Of course, thanks to my family for everything, and especially for their encouragement and moral support.
Special thanks go to Maxime who read my Master's Thesis over in order to review my English.

A huge thank you from the heart to my girlfriend for her love, patience, and encouragement.

Finally, I would like to thank all the participants who took part in the experiment. Thank you for taking a little bit of your time; it allowed me to do this work.

Antoine Moulart

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The number of Web users has doubled between 2005 and 2010, and at present exceeds two billion people [Lyn10]. A 2001 study has already estimated that "the deep Web contains nearly 550 billion individual documents compared to the one billion of the surface Web" [Ber01]. In 2011, the indexed Web contains at least 19.55 billion pages[1] and more than 131 million websites[2].

These few statistics about the World Wide Web size unquestionably show the importance of the Web today. More than these figures, we all consult websites every day to read the news, to keep in touch with friends, to buy items, to check our bank account, to book holidays, or simply to look for information.
For that purpose, we browse the web in search of a precise piece of information. Naturally, we surf the web from page to page and from site to site. Our search is thus a navigation task. "Navigation is the process through which the users achieve their purposes in using the portal or website, such as to find the information that they need or to complete the transactions that they want to do" [ZZ07].
Web Navigation is not only viewed as a process. Kalbach [Kal07] writes that "Web navigation is defined three ways:

1. The theory and practice of how people move from page to page on the Web.
2. The process of goal-directed seeking and locating hyperlinked information; browsing the Web.

[1]Maurice de Kunder, *The size of the World Wide Web*, `http://www.worldwidewebsize.com/`, (Date of access 2011-07-12).

[2]According to DomainTools, there are 131,143,510 active domains (`http://www.domaintools.com/internet-statistics/`; Date of access 2011-07-12).

3. All of the links, labels, and other elements that provide access to pages and help people orient themselves while interacting with a given website".

A lot of web designers recommend having "easy-to-navigate" sites.

In the business world area, having an effective website can improve a company's image and allow it to directly communicate with its clients [LK06]. A common characteristic of most effective sites is their high level of usability, or the user's ability to achieve his goals [NL06a]. A reason that can explain why usability is important is that the primary goal of most sites is to guide users to relevant information.

Behind the generally accepted idea that "web navigation" is important and must be well built for users, we focus in this thesis on the "navigability" issue. We use the following website navigability definition: "the ease with which users can locate and access relevant information" [VS10].

1.1 Purpose

The purpose of this Master's Thesis is to study web quality assessment. We focus on web usability and navigability assessment.
Traditionally, usability assessment is done using a (formal or informal) survey of users who assess how easy it is to perform specified tasks [ML04]. Even though these are the most complete ways to assess a system, they are also very expensive and cannot be fully integrated into a continuous development process.
An alternative to surveys is to use quality models that allow estimating how a user would react to a site. These can be used to assess quality without the prohibitive costs of actually surveying users. These models can be based on the results of a literature review or even built and calibrated using data from surveys. Their execution should be automated [IH01] so that developers can run them and get feedback quickly. It is for this purpose that we propose a fully automated approach for assessing website quality characteristics related to usability, and illustrate it on the particular case of navigability.
In concrete terms, we present and extend an existing quality model (we call it our *Multi-level Model*) that assesses website navigability thanks to a two-level assessment mechanism.

- The first level focuses on webpages and divides the navigability assessment into two separate models:

  - A **page-level model** that describes the capacity of a user to find information within a page. It computes a navigability score for a webpage.
  - A **composition model** that weights the importance of individual pages by describing how a user would navigate the site. It produces an aggregate value of quality for all pages.

- The second level focuses on the site in its entirety. The **site-level model** combines site metrics (assessment of site-level navigation elements) with the composition model output to assess the navigability of the site.

We therefore propose three "cascading models" (that form our *Multi-level Model*) to assess website navigability.
Our study extends previous work in three significant ways. First, we perform an extended assessment of the *Multi-level Model*. Second, we compare the performance of our *Multi-level Model* with the page-level model. Finally, we investigate several composition models in order to know if a certain weighting strategy improves the performance of the *Multi-level Model* or not.

1.2 Structure of the Thesis

We first review the scientific literature about website navigability (Chapter 2). This allows us to explain more accurately the context of our study. From a Software Engineering point of view, we introduce the software quality issue. We tackle the usability characteristic of software quality and the concept of "Quality Models".
Then, we focus on Web engineering. We classify research works about web quality, usability, and navigability assessment. Different approaches exist to assess website navigability. We summarize the main ones. Traditional approaches propose questionnaires, advice, guidelines, or metrics to assess website navigability. We also present some tools that are used to assess website quality.
We end this state of the art by presenting navigability assessment works according to different methods: traditional approaches (surveys and metrics), ranking algorithms, websurfer abstraction, other approaches (e.g. model-driven engineering), or a probabilistic approach.

This last category of research works is important for us because we study a quality model (our *Multi-level Model*) based on a probabilistic approach. The model initially assesses website quality characteristics related to usability. We illustrate it on the particular case of navigability.

Chapter 3 presents the *Multi-level Model* that we study.
The first version was created by **Malak et al.** [MSBB10]. They used a probabilistic approach, by means of a Bayesian Belief Network (BBN), to assess web usability. They built the model by searching a lot of web quality factors (through scientific literature about web quality) that are relevant for usability. Then, they organized them in order to have a model that allows automated website assessment. This model proceeds at the level of webpages. In the remainder of the work, we call it our "reference navigability model".
The second version of the quality model was built by **Vaucher and Sahraoui** [VS10]. They tried to improve the reference navigability model. They distinguished page-level from site-level assessment. Website assessment is always computed by means of a BBN, where the website navigability score depends on the webpages navigability score on one hand, and on the score of site-level navigation mechanisms on the other hand. These site-level navigation mechanisms are the bar menu, the intern search function, and the site map. They are assessed aside from other navigation elements. It is the *Multi-level Model* that we extend and study in this thesis.

We structure the third chapter as follows. We first define the studied problem. We introduce probabilistic quality models and the concept of Bayesian Belief Network. Then, we present our "reference navigability model" that is limited to webpages assessment. We explain why we take a different approach and focus on how to combine the individual pages assessment to the navigability assessment of the whole site.
We explain in detail our *Multi-level Model* which is composed of three models: a *"page-level model"* (based on the reference navigability model), a *"composition model"* to aggregate webpages scores and a *"site-level model"*. We explain how we try to improve the composition model by means of certain weighting strategies.

We also implement our *Multi-level Model* via a Java program. We end this third chapter by explaining the implementation structure, choices, and limits of our related tool.

Chapter 4 presents the experiment we conduct in order to assess the validity of the *Multi-level Model* presented in the previous chapter. "Validity" assessment consists of checking whether the *Multi-level Model* is able to simulate human judgment about website navigability. We follow the Wohlin and al [WRH+00] procedure about experimentation in Software Engineering.
We start by defining the object of our study, the purpose, the quality focus, and the perspective that we consider. The experiment context is also discussed. We explain in concrete terms how the experiment works. Briefly we ask our subjects to solve a task on some websites. Each task consists of searching a precise piece of information on a selected website. We then ask each subject to answer an online questionnaire in order to assess his navigation session. The subject has first to assess the navigability quality of the website on a 1 to 10 scale. We check if he succeeds in solving the task and ask him some questions about the navigation elements usefulness.
Thereafter, we define the experiment planning: research objectives and questions, variables and subjects selection, experiment design and instrumentation. The last main section of this chapter concerns the validity assessment. We review the different threats we meet and how we manage them.

Here are our two research objectives, each one being divided in two research questions.
First, we want to assess whether or not a single page-level model is able to accurately estimate human judgment. We answer the following questions: does the choice of a page impact on the result of the page-level model? Can the aggregation of individual page results correspond to human judgments about website navigability?
Second, we want to assess whether or not the *Multi-level Model* can produce better estimates. We answer two other questions: can the Multi-level Model predict human judgments about website navigability? Is there a weighting strategy that outperforms the others?

We present the experiment results in Chapter 5. To answer our research questions, we analyse the model navigability scores (computed by our Java program) and the subjects' answers to the experiment questionnaire. We compare navigability scores with human judgments and analyse the correlation coefficients. We also use statistical tests to analyse the scores distribution.
Moreover, we discuss other questionnaire results about site- and page-level navigation mechanisms: their use, their importance and how they are correlated to each other.

Based on the experiment analysis, Chapter 6 discusses some ways to improve the *Multi-level Model*. We synthesize what should be kept, changed, or introduced in our model.
Future work naturally consists of implementing and assessing our new assumptions. We introduce some new research questions too.

Chapter 2

State of the Art

This chapter describes the state of the art in assessing website quality. It lays stress on usability and navigability assessment.
First, we introduce the notion of sofware quality in the general context of Sofware Engineering. Then, we focus on web engineering and more exactly on the concept of web quality. At the core of the chapter, we review some relevant scientific papers about website usability and navigability assessment.

2.1 Software Engineering

As introduced, we remind that this Master's Thesis is about website navigability assessment. The basic frame of our study is Software Engineering.
Pressman [Pre05] uses the IEEE definition of Software Engineering:

1. "The application of a systematic, disciplined, quantifiable approach to the development, operation, and maintenance of software, and the study of these approaches; that is, the application of engineering to software.
2. The study of approaches as in (1)".

Pressman reminds us that "Software Engineering is a layered technology" (cf. Figure 2.1).
In this thesis, we focus on a particular aspect of Software Engineering, software quality. We precisely study the quality of a particular kind of software: "a website". We want to assess a particular quality that is the "usability" of a website and, more exactly, its "navigability".

Figure 2.1: Software Engineering layers (in [Pre05], page 54)

2.2 Software Quality

Software quality is "conformance to explicitly stated functional and performance requirements, explicitly documented development standards, and implicit characteristics that are excpected of all professionnally developed software" [Pre05].
Pressman [Pre05] speaks about two kinds of quality: quality of design (refers to the characteristics that designers specify for an item) and quality of conformance (is the degree to which the design specifications are followed during manufacturing). He adds that software engineers must consider an additional issue: "user satisfaction".

We can also quote different definitions of **software quality**.
Software Quality [definition from IEEE-Std-729]

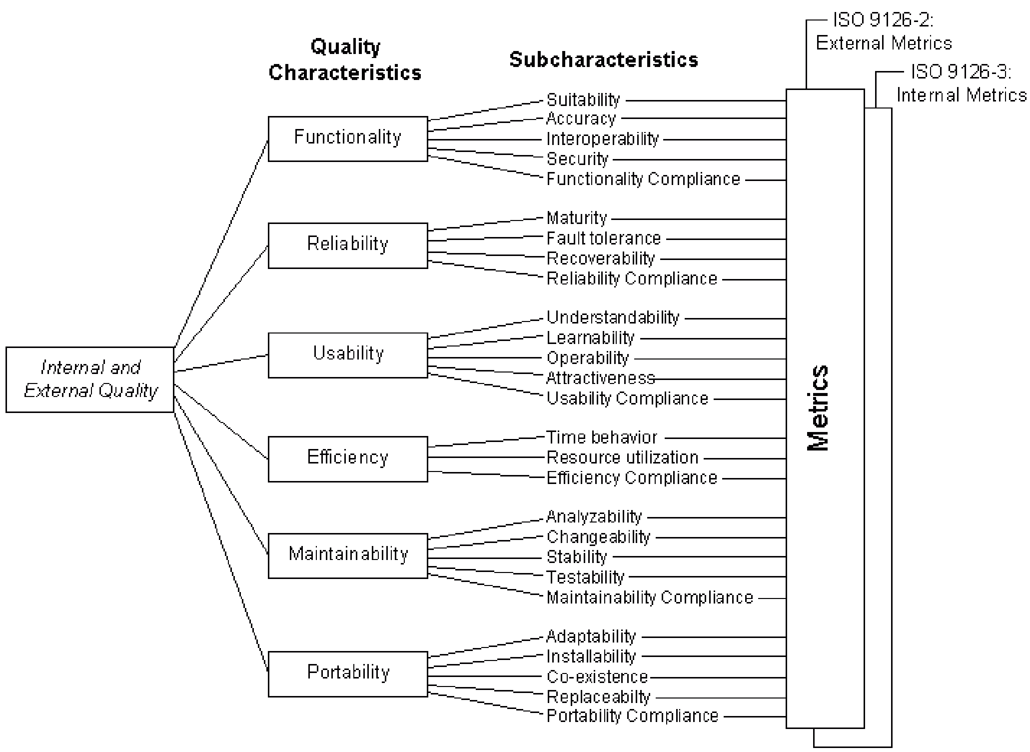
- "The totality of features and characteristics of a software product that bear on its ability to satisfy given needs.
- The degree to which software possesses a desired combination of attributes.
- The degree to which a customer or a user perceives that a software meets her composite expectations.
- The composite characteristics of a software that determine the degree to which the software in use will meet the expectations of the customer".

Software quality characteristics [definition from ISO-9126] are "a set of software product attributes by which its quality is described and evaluated".
The ISO 9126 definition defines six characteristics and a set of sub-characteristics for software quality. Figure 2.2 illustrates this definition[1].

[1]Picture from `http://www.chrisbunney.com/wiki/index.php/File:ISO9126.gif`; (Date of access 2011-02-05).

- "Functionnality: The capability of the software product to provide functions which meet *stated and implied* needs when the software is used under specified conditions.
- Reliability: The capability of the software product to maintain a specified level of performance when used under specified conditions.
- Usability: The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions.
- Efficiency: The capability of the software product to provide appropriate performance, relative to the amount of resources used, under stated conditions.
- Maintainability: The capability of the software product to be modified. Modifications may include corrections, improvements, or adaptation of the software to changes in environment, and in requirements and functional specifications.
- Portability: The capability of the software product to be transferred from one environment to another".

Figure 2.2: Software Quality Definition from ISO9126.



Bevan [Bev99] follows ISO9126 and reminds us of three perspectives on software quality: "internal quality (static properties of the code), external quality (behaviour of the software when it is executed), and quality in use (whether the software meets the needs of the user when it is in use)".

Fenton and Neil [FN00] study software metrics, i.e. how to associate numeric value to quality characteristics. They recommend "to handle the key factors largely missing from the usual metrics approaches, namely: causality, uncertainty, and combining different (often subjective) evidence". They propose using "causal modelling (using Bayesian nets), empirical software engineering, and multi-criteria decision aids".

We now approach the usability characteristic of this definition.

2.2.1 Software Usability

We focus our attention on the usability characteristic of the software quality definition from ISO9126. Here are three definitions of usability from the ISO/IEC 9126-1 quality model:

- "A set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users.
- The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.
- The capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions".

The usability sub-characteristics are the following:

- "Understandability: capability to enable the user to understand whether the software is suitable, and how it can be used for particular tasks and conditions of use.
- Learnability: capability to enable the user to learn its application.
- Operability: capability to enable the user to operate and control it.
- Attractiveness: capability to be attractive to the user.

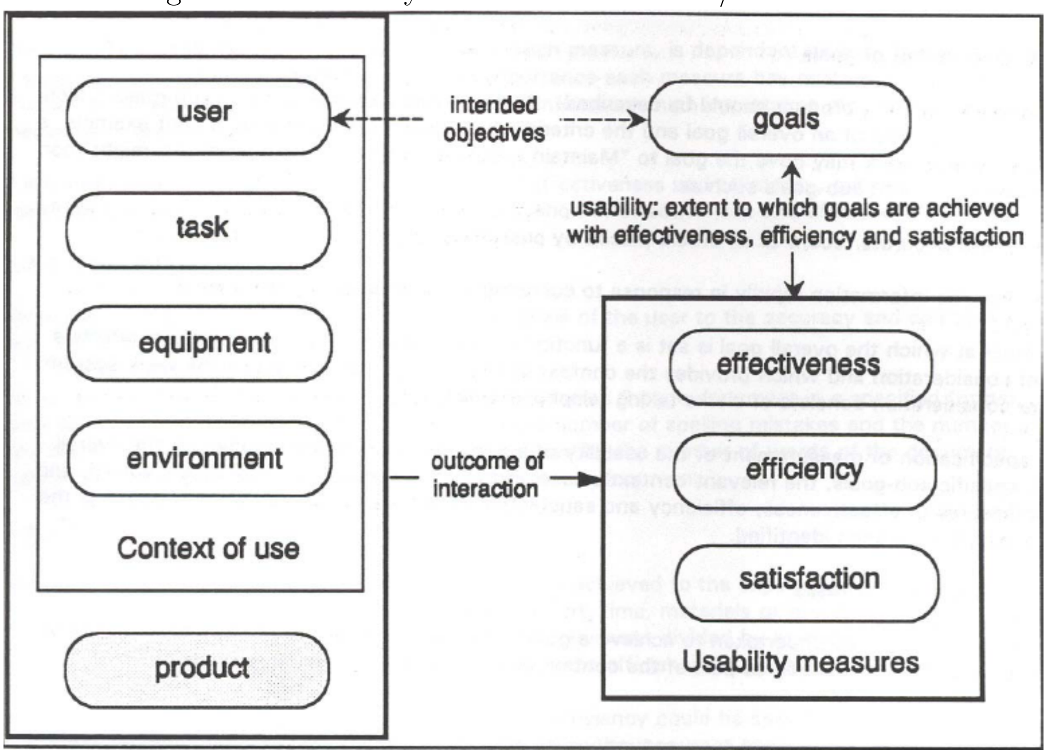
- Usability compliance: capability to adhere to standards, conventions, style guides or regulations relating to usability".

"ISO 9241-11 explains the benefits of measuring usability in terms of user performance and satisfaction". It is defined by three sub-characteristics:

- "Effectiveness (the extent to which the intended goals of use are achieved),
- Efficiency (the resources that have to be expended to achieve the intended goals), and
- Satisfaction (the extent to which the user finds the use of the product acceptable)".

Figure 2.3 presents the usability framework according to ISO9241-11.

Figure 2.3: Usability Framework from ISO/DIS 9241-11.



These definitions concern common softwares. In this work, we focus on a particular kind of software: "websites". In the context of website, we present an adapted definition of the "usability" characteristic. Then, we speak about "navigability". Before tackling these topics, we introduce the "Quality Model" notion.

2.2.2 Software Quality Models

According to ISO/IEC 25000:2005, a quality model (QM) is a "defined set of characteristics, and of relationships between them, which provides a framework for specifying quality requirements and evaluating quality". "The goal of a quality model is in essence to provide an operational definition of quality"[aCMIPG]. "A formal framework must precisely and unambiguously provide meanings for common concepts and terms and do so in a readable and understandable manner" [Ld01].

Several frameworks and quality models have been developed to assess Sofware Quality. For example, **Vanderose, Kamseu, and Habra** [VHK10] present "MoCQA, a Model-Centric Quality Assessment framework relying on a quality metamodel and supporting a flexible integration of different types of quantitative quality assessment all along the software development lifecycle". They use a two-level methodology: firstly, the framework generates a customised assessment quality model and secondly, it implements a concrete measurement plan.

We have thus introduced the notion of quality in Software Engineering, and emphasized the "usability" characteristic. Quality Models are used to detail quality characteristics (and their relationships) but also to assess software quality. In the remainder of this chapter, we focus on the Web context by introducing Web Engineering. We then discourse on web quality, website usability and navigability, and their assessment.

2.3 Web engineering

Common softwares are the subject of study of Software Engineering. Websites are particular software products which represent the subject of study of a more specialized domain called "Web Engineering". In this section, we tackle Web Engineering before providing a literature review about web quality assessment .

Web engineering [MDHG01] is "the establishment and use of sound scientific, engineering and management principles and disciplined and systematic approaches to the successful development, deployment and maintenance of high quality Web-based systems and applications".

Web Engineering[2] is also defined as "the application of systematic, disciplined and quantifiable approaches to the cost-effective development and evolution of high-quality solutions in the World Wide Web".

Stephens [Ste01] drew up the Web Engineering literature review. **Suh** [Suh05] defined Web engineering as "the way of developing and organising knowledge about Web application development and applying that knowledge to develop Web applications, or to address new requirements or challenges. It is also a way of managing the complexity and diversity of Web applications. Web engineering is multidisciplinary and encompasses contributions from diverse areas :

- systems analysis and design,
- software engineering,
- hypermedia/hypertext engineering,
- requirements engineering,
- human-computer interaction,
- user interface,
- information engineering,
- information indexing and retrieval,
- testing,
- modelling and simulation,
- project management, and
- graphic design and presentation".

According to these definitions, we notice that Web Engineering is not only a Software Engineering subset. Web Engineering focuses on approaches, methods, models, principles and tools that are specific to web-based applications[3]. Keeping in mind this Web Engineering context, we focus now on websites and on the "web quality" notion.

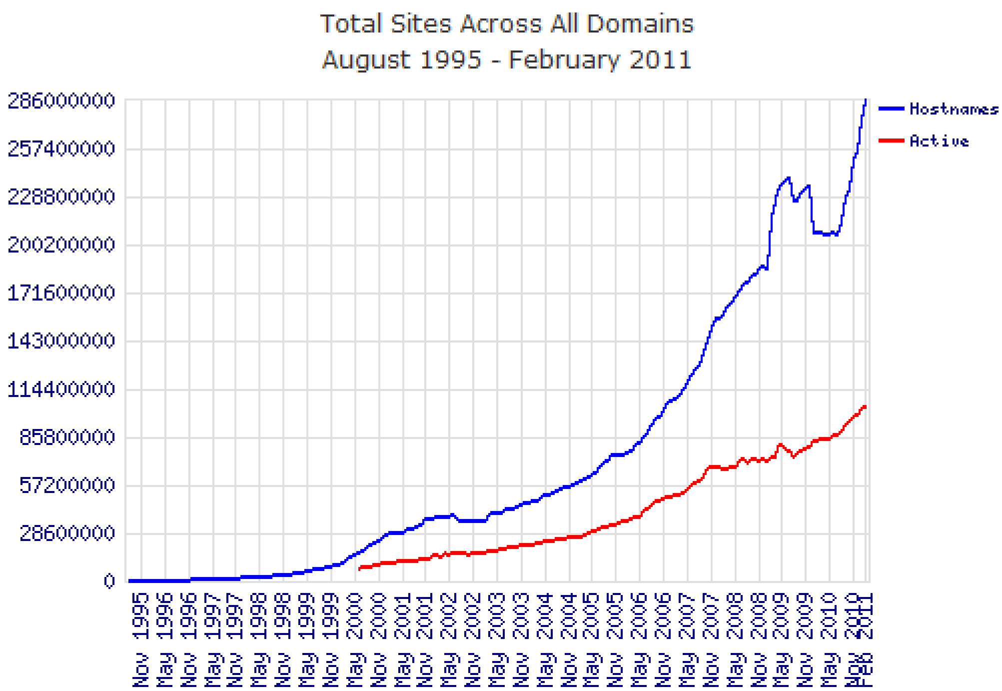
[2] Definition from `http://www.webengineering.org/`; (Date of access 2011-02-05).
[3] According to `http://www.iswe-ev.de/`, International Society for Web Engineering; (Date of access 2011-02-05).

2.4 Web Quality

Websites are particular software applications hugely used by a wide public. For a lot of companies, websites are now the business basis. As for the general public, the web is today ubiquitous. It allows the user to carry out lots of different tasks such as searching a piece of information, booking holidays, sharing multimedia documents, communicating with friends, etc.

In February 2011, Netcraft[4], an Internet monitoring company that has tracked

Figure 2.4: Websites study from Netcraft [February 2011]



Web growth since 1995, reported that they "received responses from 284,842,077 sites". The Web has thus become hugely used. In order to provide services that are easy to use for Web surfers, websites need to have a certain quality. In the following paragraphs, we define Web quality and summarize some works, tools, and models about Web quality. Further, we approach other works, papers, or tools concerning more accurately Web usability and navigability.

[4]Posted by Jennifer Cownie on 15th February, 2011 ; `http://news.netcraft.com/archives/2011/02/15/february-2011-web-server-survey.html`

2.4.1 Web Quality Definition

In fact, it is quite difficult to give a consensual definition of "Web Quality". Indeed, the Web Engineering is a relatively new and multi-disciplinary domain. Some researchers have based their studies on the ISO definiton of "Software Quality" to explain what "Web Quality" exactly is. In the following paragraphs, we summarize some works and studies about "Web Quality". Despite the lack of a worldwide recognized definition, we can introduce two attempts of "Web Quality" definition :

Firstly, **a well-engineered Web system** [Suh05] is defined as:

- "Functionally complete and correct;
- Usable;
- Robust and reliable;
- Maintainable;
- Secure;
- Perform satisfactory even under flash and peak loads;
- Scalable;
- Portable, where required perform across different common platforms; compatible with multiple browsers;
- Reusable;
- Interoperable with other Web and information systems;
- Universal accessibility (access by people with different kinds disabilities);
- Well-documented".

Secondly, we can add a brief description of the "Quality control manager" job[5] that is emerging today at the web level. Malassingne[6] gives the following definitions:

Web Quality is quality norms (objective and measurable) management in order to meet a defined goal.

[5]French term "qualiticien".

[6]Delphine Malassingne, *Mettre en place la gestion de la qualité web*, on `http://articles.nissone.com/2010/10/mettre-en-place-la-gestion-de-la-qualite-web/`, 19/10/2010 (Date of access 2011-02-12).

Web quality control manager: A "Web Quality control manager" has a support function. He works with the production teams, process, and tools, in order to help them work better and meet the defined goals.

2.4.2 Web Quality Standards

Two main organizations provide Web Quality standards: W3C and IEEE.

The World Wide Web Consortium (W3C) develops Web standards (Web design and applications standards for building and rendering Web pages as HTML, CSS, SVG, Ajax; Web architecture standards including URIs and HTTP, Semantic Web standards, etc.). W3C reduces "Web Quality" to W3C standards validation. For example, the "W3C Quality Assurance Interest Group" puts forward techniques and tools to ensure Web standards validation[7].

[IEE03] gives best practices to improve "the productivity of intranet/extranet Web operations in terms of locating relevant information, and efficient development and maintenance practices". It tackles design practices, server and HTTP considerations, header information, and body information. It recommends practices for well-engineered webpages, based on HTML specifications, migration to XML (according to the W3C), and related industry guidelines.

Other web quality standards exist, such as OPQUAST (Open Quality Standards) that identifies 217 best practices in order to improve web quality[8].

2.4.3 Web Quality Research Trends

We now summarize the main ideas of previous work on web quality. Studies that focus on web usability and navigability are presented in the next sections.
More information about the state of the art in "Web-based applications Quality Assessment" can be found in [BMBB02], where **Malak et al.** proposed "a survey in the field of quality insurance within the framework of web based applications by evaluating proposed approaches, criterions and metrics".

[7]From `http://www.w3.org/standards/` and `http://www.w3.org/QA/2002/04/Web-Quality`; (Date of access 2011-02-12).

[8]From `http://checklists.opquast.com/opquastv2`; (Date of access 2011-02-12).

The Hypertext Quality

First research trends on web quality focused on the hypertext, and especially on its structure. **Brown** [Bro90] explained how to write hyperdocuments "that have long lifetime, can be maintained and have a 'correct' structure according to certain rules". **Botafogo et al.** [BRS92] used graph theory and other Software Engineering methods to identify concepts linked to coherence and coupling in the hypertext. They proposed several tools, based on hypertext structure analysis, to solve the "lost in hyperspace" problem. They introduced metrics to compare different hypertexts and to study their properties. They noticed that "interpretations made from metrics should be carefully verified". Similarly, **Hatzimanikatis et al** [HTC95] defined a hyperdocument quality model. They presented "structure metrics" (derived from well-known software metrics). They especially measured the hypertext readability and usability.

Then, researchers studied more than the structure of the hypertext. For example, **Garzotto et al.** [GMP95] "identified several dimensions for analyzing a hypermedia application: content, structure, presentation, dynamics, and interaction". They also suggested other quality criteria for a hypertext application: information richness, ease of accessibility, consistency, self-evidence, predictibility, readability and reuse. They applied their "design-oriented evaluation method" to a popular commercial application (a hypermedia guide).

The Web: Websites Quality

Thereafter, researchers focused on the web, web applications, and websites. **Bray** [Bra96] studied general questions about the web. He tried "to provide partial qualitative answers to questions like how big is the web, whats an average page like, what are the most visible websites, how richly is the web connected and what data formats are being used". He introduced the measure of "HTML sincerity" and showed that "a large majority of pages (over 87%) are making some effort to present themselves as HTML". Bray is the first to study webpages quality.

Then, researchers studied the measurement of websites (establishing metrics and attributes). They based their assessment on software engineering metrics. For example, **Boldyreff et al.** [BW00] explained that the first thing to do was "identify consistent ways of measuring the success of a site as a whole". They proposed "a scientific approach in order to establish metrics, based in part on current software engineering metrics research". **Offutt** [Off02] discussed "some of the technological challenges of building today's complex Web software applications, their unique quality requirements, and how to achieve them". He surveyed and discussed the important quality process drivers for Web applications. Web software development managers and practitioners considered that the three most important quality criteria were Reliability, Usability, and Security. Additional important criteria include Availability, Scalability, Maintainability, and Time to market.

As introduced, **Malak et al.** [BMBB02] wrote the state of the art in "Web-based applications Quality Assessment". They found 290 different criteria and sub-criteria and developped a checklist and a "quality tree" for Web Applications assessment. Then, they proposed a method for building web application quality models by means of Bayesian networks.

Consideration of the User

Some studies have emphasized that users or clients have to be taken into account in order to assess website quality. For example, **Barnes and Vidgen** [BV00] used the 'voice of the customer' in order to tackle the issue of website quality. They presented "a framework for identifying web-site qualities demanded by users, which were gathered through a quality workshop. From the workshop an instrument for assessing web-site quality was developed (WebQual) and tested in the domain of UK business schools". **Cowderoy** [Cow00] studied the context of website development and explained that "a distinction must be maintained between internal and external measures, with aggregate measures and cost models separated from these". About website specifications, he said that "often the only way to achieve al good site was by repeated experimentation and by establishing a very close working relationship between the client and the creative people in the multimedia developers". **Huang** [Hua03] identified "Web attributes, their direct impacts on experiential flow, and their direct and indirect impacts on the utilitarian and hedonic aspects of Web performance". He concluded that "a successful website must be able to use its attributes to satisfy both the information and entertainment needs of users. **Ziemer and Stalhane** [ZS06] "took a special interest into how quality issues are managed in web application development". They found that this kind of development was "communication intensive" (with users). In their study about companies developing web applications, they discovered that "the most important quality factors mentioned were availability and reliability, performance and to give the users a good user experience".

Specific Areas

Several studies about Web quality assessment focused on specific areas, especially on e-commerce websites. For example, **Shubert et al.** [SD02] developed EWAM (Extended Web Assessment Method). It is applied to e-commerce websites. "It defines an evaluation grid including a set of criteria to appraise the quality and success of existing e-commerce applications". **Albuquerque** [AB02] proposed the FMSQE model (Fuzzy Model for Software Quality Evaluation). This model uses fuzzy logic. "It identifies and ranks the main quality attributes to the application domain of e-commerce websites". **Lohse and Spiller** [LS98] "surveyed 35 attributes of 137 Internet retail stores to provide a classification of the strategies pursued in Web-based marketing". They also "measured 32 interface features for 28 online retail stores in August 1996 and identified store design features that influence online store traffic and sales". **Stefani and Xenos** [SX08] presented a model for the quality of e-commerce systems, "based on Bayesian Networks and ISO 9126. Besides the emphasis on specific software quality attributes, it also provides a quality assessment process aiding developers to design and produce e-commerce systems of high quality". **Webb et al.** [WW04] focused on B2C Web site quality. They developed SiteQual, "a conceptual model and an instrument to measure Web site quality". They conducted a factor analysis to find quality factors that "are important to consumers in the retail music industry". They suggested "the use of Website quality factors for measurement of consumer expectations and perceptions, determining Website requirements, and guiding the testing process".

Some researchers discussed other specific areas. For example, **Olsina et al.** [OLR01] studied the quality of academic websites. They found more than a hundred characteristics and attributes and they built a Quality Requirement Tree. They proposed WebQEM (Web Quality Evaluation Methodology). This approach aims at assessing "the artifact quality in the operational phase of a Web Information System (WIS) lifecycle". **Kedowide** [Ked08] proposed an adaptation of the WebQEM method to assess the quality of the Canadian Human Resources Minister website.

2.4.4 Web Quality Models

Quality models are initially used to assess software quality (cf. *Software Quality* section; *Sotware Quality Models*). Researchers developed specific quality models to assess web quality, largely inspired by software quality models.

We present such web quality models. **Ruiz et al.** [RCP03] proposed a model in order to assess web quality. "The model can be used for the classification of web metrics and web research works". The three dimensions of the model are: the quality characteristics dimension (based on the ISO 9126 standard), the life cycle process dimension, and the website features dimension (contents, functions, infrastructure, environment). **Mich et al.** presented the 2QCV3Q[9] model [MFC03] ; [MFG03]. "The 2QCV3Q model helps developers evaluate Website quality from both owner and user viewpoints. The 2QCV3Q model provides a conceptual framework for identifying aspects that determine overall Website quality". **Mavromoustakos et al.** [MA07] proposed the WAQE (Web Application Quality Evaluation) model. It is based on an internal (the organisation) and an external axon (the users). They used the ISO 9126 quality issues and other web quality factors. They developed "importance-based criteria for evaluating requirements". **Caro et al.** [CCdSP07] presented PDQM (Portal Data Quality Model). They identified 33 data quality attributes for the theoretical version. Then, they "adopted a probabilistic approach by using Bayesian networks" in order to create the operational model. PDQM has been implemented in the PoDQA (Portal Data Quality Assesment) tool[10].

2.4.5 Web Applications Assessment Tools

Several tools have been developed to assess the quality of web applications, based on ergonomic recommendations, best practices, or standards.

[9] Dimensions of the 2QCV3Q model: QVIS? (Who) → Identity ; QVID? (What) → Content; CVR? (Why) → Services; VBI? (Where) → Location; QVANDO? (When) → Maintenance; QVOMODO? (How) → Usability; QVIBVS AVXILIIS? (With what means and devices) → Feasibility.

[10] `http://podqa.webportalquality.com/`; (Date of access 2011-02-12).

For example, Watchfire/WebXM[11] is a Web quality monitoring and reporting tool that scans websites and reports “issues such as spelling errors, broken links, and outdated content”. **Lanzi et al.** [LMM04] proposed a “method and a toolset for quality evaluation of Web applications that exploits conceptual specifications, deriving from the adoption of model-based development methods, for the evaluation in pre- and post- delivery phases.” Their framework supports three kinds of analysis: “the Design Schema Analysis (DSA) verifies the correctness and the internal coherence of specifications” ; “Web Usage Analysis (WUA) produces reports on content access and navigation paths followed by users ; Web Usage Mining (WUM) applies XML mining techniques for discovering interesting (sometimes unexpected) associations between accessed data”. **Guillemot and König** [GK06] described WebTest: “an Open Source tool for automated testing of web applications”. **Ricca and Tonella** [RT06] studied 15 analysis and testing tools for Web applications. They concluded that “only research prototypes, such as ReWeb and TestWeb, offer the most advanced features (such as reverse engineering of high-level models and structure-based testing) in order to discover many anomalies and failures in Web applications”.

2.5 Web Usability

For general software, usability is one of the main characteristics of sofware quality, as previously introduced (cf. *“Software Quality”* section; *“Software Usability”*). **Web usability** is the application of (general software) usability to web applications.

According to *Usability.gov*[12], “Usability measures the quality of a user’s experience when interacting with a product or system - whether a website, a software application, mobile technology, or any user-operated device. It is important to realize that usability is not a single, one-dimensional property of a user interface. Usability is a combination of factors including:

- Ease of learning - How fast can a user who has never seen the user interface before learn it sufficiently well to accomplish basic tasks?
- Efficiency of use - Once an experienced user has learned to use the system, how fast can he or she accomplish tasks?

[11] `http://www.utexas.edu/its/watchfire/`; (Date of access 2011-02-12).
[12] `http://www.usability.gov/basics/index.html`; (Date of access 2011-02-13).

- Memorability - If a user has used the system before, can he or she remember enough to use it effectively the next time or does the user have to start over again learning everything?
- Error frequency and severity - How often do users make errors while using the system, how serious are these errors, and how do users recover from these errors?
- Subjective satisfaction - How much does the user like using the system?"

This usability definition comes from **Nielsen**[13]: "Usability is a quality attribute that assesses how easy user interfaces are to use. The word *usability* also refers to methods for improving ease-of-use during the design process. Usability is defined by 5 quality components: Learnability, Efficiency, Memorability, Errors, and Satisfaction". Nielsen explained why website usability is important: "On the Web, usability is a necessary condition for survival. If a website is difficult to use, people leave. If the homepage fails to clearly state what a company offers and what users can do on the site, people leave. If users get lost on a website, they leave. If a website's information is hard to read or doesn't answer users' key questions, they leave. Note a pattern here? There's no such thing as a user reading a website manual or otherwise spending much time trying to figure out an interface. There are plenty of other websites available; leaving is the first line of defense when users encounter a difficulty".

Finally, we can find several "usability" definitions on the website of the Usability Professionals' Association[14].

2.5.1 Research Trends about Web Usability Assessment

We summarize the main research trends about web usability assessment. More information can be found in [IH01], where **Ivory and Hearst** drew up the state of the art in automating usability evaluation of user interfaces. They presented "an extensive survey of usability evaluation methods, organized according to a new taxonomy that emphasizes the role of automation. The survey analyzes existing techniques, identifies which aspects of usability evaluation automation are

[13] According to Jakob Nielsen, *Usability 101: Introduction to Usability*, on `http://www.useit.com/alertbox/20030825.html`; (Date of access 2011-04-12).

[14] `http://www.usabilityprofessionals.org/usability_resources/about_usability/definitions.html`; (Date of access 2011-03-08).

likely to be of use in future research, and suggests new ways to expand existing approaches to better support usability evaluation".

Scientific literature often refers to **Nielsen**, who defined "nine usability heuristics [NM90]:

- Use simple and natural dialog
- Speak the user's language
- Minimize the user's memory load
- Be consistent
- Provide feedback
- Provide clearly marked exits
- Provide shortcuts
- Prevent errors
- Provide helpful error messages".

Nielsen focused on "Usability Engineering". He especially proposed the heuristic evaluation, which "is an informal method of usability analysis where a number of evaluators are presented with an interface design and asked to comment on it" [NM90]. He also reviewed 8 types of usability inspection methods: heuristic evaluation, cognitive walkthroughs, formal usability inspections, pluralistic walkthroughs, feature inspection, consistency inspection, and standards inspection [Nie95]. **Nielsen and Loranger** [NL06b] gave advice and guidelines in order to address Web Usability issues. Many web usability guidelines can be found on the *Nielsen Norman Group* website[15].

Shum et al. [SM97] introduced the web usability issue because they "felt there was something missing between the vast amount of hypermedia and related human-computer interaction (HCI) research that has been conducted, and the most popular hypermedia system in existence: the World Wide Web". **Morville**[16] studied "the relationship between information architecture design and usability engineering".

[15] `http://www.nngroup.com/reports/`; (Date of access 2011-03-06).

[16] Peter Morville ; *Information, Architecture, and Usability*; December 4, 2009 ; from `http://semanticstudios.com/publications/web_architect/usability.html`; (Date of access 2011-04-21).

A User-centered Approach

The first usability assessment methods consisted mainly of questionnaires or user surveys. For example, **Mehlenbacher** [Meh93] identified 8 methods to assess online documents and systems (with the advantages and drawbacks of each usability method). The usability test methods were "Talk-Aloud Protocols, Videotaped Sessions, Interviews, User Surveys, System Benchmarking, The Wizard of Oz Technique, Guided Interaction, and Beta-Testing". **Kirakowski et al.** [KC98] developed the WAMMI questionnaire. This form is based on user satisfaction. It provides developers with a list of elements to diagnose usability problems and improve website quality. Today, the WAMMI questionnaire is available as a paid service[17]. **De Marsico and Levialdi** [ML04] exploited users' expectations in order to create a new goal-based approach to measure websites usability.

Checklists and Metrics

Several checklists, guidelines, and metrics were developed to automate the web usability assessment.
Keevil [Kee98] developed a checklist to measure a website usability index (that is "a measure, expressed as a per cent, of how closely the features of a website match generally accepted usability guidelines"). An online version of the checklist exists[18]. **Lowe and Hall** [LH99] proposed metrics directly linked to the usability: navigability, links validity, and organization of hypermedia applications. They were interested in the size and life time of these applications.
Palmer [Pal02] led a series of three studies that developed and validated Website usability, design and performance metrics, including download delay, navigability, site content, interactivity, and responsiveness. The performance metric that was developped includes the subconstructs user satisfaction, the likelihood of return, and the frequency of use. Palmer gave five elements associated with users being-are satisfied by websites: websites exhibiting lower download delay; more navigable websites; higher interactivity in websites; more responsive websites; and higher quality content in websites. **HHS** (U.S. Department of Health and

[17]WAMMI questionnaire, on `http://www.wammi.com/index.html`; (Date of access 2011-02-25).

[18]Measuring the Usability of Your Web Site, by Benjamin Keevil, on `http://www3.sympatico.ca/bkeevil/sigdoc98/checklist/WebCheck_Sep13.html`; (Date of access 2011-04-21).

Human Services) provides a complete online guide[19] that gathers and organizes 209 usability guidelines.

Methods and Models

Kantner et al. [KSR02] presented “a structured process for evaluating the usability of online documentation, based on a list of heuristics for navigating through and finding content”. **Kitajima et al.** [KKTZ05] presented “a method to quantitatively evaluate the usability of large-scale information-oriented websites and the effects of improvements made to the site design. This was achieved by utilizing the Cognitive Walkthrough for the Web and website modeling using Markov chains”. They “demonstrated that the average number of clicks before a visitor reaches a goal can be analyzed simply and that the effect of the usability improvement method suggested by the Cognitive Walkthrough for the Web can be evaluated quantitatively”. **Seffah et al.** [SDKP06] presented “a single consolidated, hierarchical model of usability measurement, called Quality in Use Integrated Measurement (QUIM). Included in the QUIM model are 10 factors each of which corresponds to a specific facet of usability that is identified in an existing standard or model. These 10 factors are decomposed into a total of 26 sub-factors or measurable criteria that are furtherdecomposed into 127 specific metrics”. **Frokjaer and Hornbaek** [FH08] presented a new technique (MOT) that guides inspection by metaphors of human thinking. They led three experiments and concluded that “usability problems uncovered with MOT were more serious and more complex to repair than problems found with heuristic evaluation. MOT found also more problems than cognitive walkthrough, and has a wider coverage of a reference collection of usability problems”.

Accessibility

“Universal usability[20] accounts for users of all ages, experience levels, and physical or sensory limitations”. Several standards and guidelines especially focus on the **accessibility** in order to achieve Universal Usability. In this work, we do not focus on the accessibility issue, which we only introduce.

“Usability pertains to the layout, the location of elements, the functionality of the progressive enhancements, the design, the site’s inherit intuitiveness, and more.

[19] `http://www.usability.gov/guidelines/`; (Date of access 2011-07-13).
[20] `http://webstyleguide.com/wsg3/2-universal-usability/4-guidelines.html`

Simply put, **web accessibility** is the ability to access web content"[21].
Whitelaw [Whi03] discussed "six things common to web pages that can easily be made more accessible: Images, Links, Color, Tables, Headings, and Navigation".
W3C's Web Accessibility Initiative (WAI) has published Web Content Accessibility Guidelines (WCAG) to help authors create content that is accessible to people with disabilities. The W3C explains that the *ISO 9241 Usability definition* is close to its goal with WAI: "the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments"[22].
A complete list of Web Accessibility Evaluation Tools is available on the W3C website[23]. Another Web accessibility checklist is proposed by IBM[24].

2.5.2 Tools to assess web usability

Several tools exist to assess web usability. For example, **Ivory and Hearst** [IH02] proposed a tool (Quality Checkers for Web Site Designs) that partially automates the usability evaluation process. They "developed a system that computes a set of quantitative measures that can characterize the informational, navigational, and graphical aspects of the web site, and have shown that even a small set of such measures can be used to successfully distinguish highly-rated web sites from poorly-rated ones". **Alva et al.** [OPL$^+$03] presented "the evaluation of a group of methods and tools for the measurement of usability in software products and software artefacts in the web". **Matera et al.** [MRTC06] introduced "principles and evaluation methods to be adopted during the whole application lifecycle for promoting usability". **Mariage et al.** [MVBM04] developed the DESTINE tool, which is "conceived in order to evaluate the usability of websites. Evaluation is made automatically for a set of guidelines, and manually for non automatisable guidelines. DESTINE covers some steps of the evaluation task, from the context specification to the finalization and the communication of the evaluation report."

[21] `http://accessites.org/site/2007/09/a-comparative-accessibility-and-usability/`; (Date of access 2011-07-08).
[22] `http://www.w3.org/2010/11/dd-wud.html#%283%29`; (Date of access 2011-07-09).
[23] `http://www.w3.org/WAI/ER/tools/complete.html`; (Date of access 2011-07-09).
[24] `http://www-03.ibm.com/able/guidelines/web/accessweb.html`; (Date of access 2011-07-09).

2.6 Website navigability assessment

Malak [Mal07] studied the navigability issue and explained that "the navigability design concerns in fact three quality characteristics: usability, fonctionnality, and efficiency". However, Malak noticed that most researchers (e.g. Nielsen, Ivory, Koyani) tend to say that "the navigability design characterizes the usability of a web application".

The main definition of navigability comes from **Palmer** [Pal02]. He wrote: "Navigation is an important design element, allowing users to acquire more of the information they are seeking and making the information easier to find. Thus, a key challenge in building a usable Web site is to create good links and navigation mechanisms. Graphical design, layout, and actual content are prime components in making the page easier to use. Text links are vital; navigation and content are inseparable; and key areas are navigational structure, searching, readability, and graphics". In Palmer's opinion, navigability is defined as "the sequencing of pages, well organized layout, and consistency of navigation protocols".

Following **Vaucher et al.** [VS10], we define navigability as follows: "The navigability of a web site is a measure of how easily a user can locate and access the information he needs". Our definition is very similar to the **Zhou**'s definition [ZLW07]: "Navigability denotes the ease with which users can find a required piece of information as they move from a home- page and on through a Web site".

We now present five families of approaches that aim at assessing website navigability: traditional approaches and development of metrics, ranking algorithms, websurfer abstraction, other approaches, and probabilistic approaches based on Bayesian Networks.

2.6.1 Traditional Approaches & Metrics

We have already introduced several traditional approaches that focus on the hypertext quality (cf. *Web Quality* section; *Web Quality Research Trends*). **Zhou et al.** [ZLW07] remind us that "existing navigability measures are based mainly on the static hyperlink structure of a website. The two important findings in previous empirical studies on navigability are that hypertexts with a moderate

level of breadth and depth actually afford optimal navigation performance; and hypertexts with too few or too many cross-referential hyperlinks imply low navigability".
Other researchers proposed metrics in order to measure website navigability. For example, **Zhang et al.** [ZZG04] proposed five metrics for website navigability measurement, based on the website structural complexity. Their empirical study shows that structural complexity plays a significant role in Web navigability. Hence, website structural complexity metrics can be used to measure web navigability indirectly.

Sreedhar et al. [SVC10] measured the quality of website navigation. They found key components: "the sitemap, a path length metric (used to evaluate average number of clicks to get the desired webpage) and the website structural complexity (determined with cyclomatic complexity)". **Bonsón-Ponte et al** [BPCGER08] analysed the navigation quality of the websites of Lithuanian banks. "Each website was rated on the basis of a series of metrics that measure the quality of navigation, in accordance with the Web Quality Model developed by Calero et al. [RCP03]".

2.6.2 Ranking algorithms

Researchers studied ranking algorithms, for example to assess which webpages are the most important on a website. **White et al.** [WS03] focused on "defining and computing the importance of nodes in a graph relative to one or more root nodes". They defined "a general framework and a number of different algorithms, building on ideas from social networks, graph theory, Markov models, and Web graph analysis". **Zhang et al.** [ZL06] proposed a novel ranking algorithm called XRank as a solution to three problems met with PageRank (the most famous link analysis algorithm that offers an effective way to rank the pages). **Diligenti et al.** [DGM04] proposed "a general probabilistic framework for Web Page Scoring Systems (WPSS), which incorporates and extends many of the relevant models proposed in the literature". They introduced "scoring systems for both generic (horizontal) and focused (vertical) search engines. Whereas horizontal scoring algorithms are only based on the topology of the Web graph, vertical ranking also takes the page contents into account and are the base for focused and user adapted search interfaces".

2.6.3 Websurfer Abstraction

Several researchers proposed navigability assessment methods that model the actions of a web surfer.
Pirolli and Fu [PF03] developed a computational cognitive model called SNIF-ACT (Scent-based Navigation and Information Foraging in the ACT cognitive architecture). It explains navigation behaviour on the Web. Based on the Information Foraging Theory (IFT), the authors used a spreading activation mechanism to quantify the perceived relevance of a Web link to a user's goal. "SNIF-ACT 1.0 utilizes the measure of utility, called *information scent*, derived from IFT to predict rankings of links on different Web pages". **Chi et al.** [CRS+03] described "a prototype service called InfoScent Bloodhound Simulator, a push-button navigation analysis system, which automatically analyzes the information cues on a Website to produce a usability report". They built "a method called Information Scent Absorption Rate, which measures the navigability of a site by computing the probability of users reaching the desired destinations on the site".
Katsanos et al. [KTA06] presented InfoScent Evaluator, "a tool that automatically evaluates the semantic appropriateness of the descriptions of hyperlinks in web pages". They studied "how information scent[25], this important attribute of hypermedia navigability, influences concurrently four aspects of users' behaviour while exploring a website: distribution of attention; confidence in choice of link; efficiency; and effectiveness. The websites were evaluated through eye-tracking user studies" [KTA10].

Diligenti et al. [DGM04] defined "a general probabilistic framework for random walks". They modeled the actions of a generic Web surfer by a set of conditional probabilities which depend on the current page q (the probability of following a hyperlink from q; the probability of following a back-link from q; the probability of jumping from q; the probability of remaining q). Similarly, **Zhou et al.** [ZLW07] introduced a navigability measure called MNav. It is based on the abstraction of "a dynamic Web surfing behavior as a Markov model which synthesizes typical surfing actions". The five actions are "terminating-session"; "proceeding-to" (follow a hyperlink); "going-back"; "staying-in" and "jumping-

[25] Information "scent": a user's "(imperfect) perception of the value, cost, or access path of information sources obtained from proximal cues, such as WWW links" [KTA06].

to". They showed that "MNav could be efficiently computed and it provided an effective and useful measurement of website navigability".

2.6.4 Other Approaches

Here are other research that assess the navigability differently.
Fang et al. [FCH+06] proposed "a systematic website navigability evaluation method built on Web mining techniques". They developed "three objective metrics for measuring website navigability on the basis of the Law of Surfing". **Wheeldon and Levene** [WL03] presented "an algorithm called the Best Trail Algorithm, which helps solve the hypertext navigation problem by automating the construction of memex-like trails[26] through the corpus."

Cachero et al. [CMG+07] presented a Model-Driven Engineering approach. This generic approach is used to define "navigability measurement models that can be integrated into a Web engineering methodology".

2.6.5 Probabilistic Approaches based on Bayesian Networks

Website navigability can be assessed by means of a probabilistic approach based on Bayesian Networks. Here are the main studies that we follow in this work.
Haydar et al. [HMS+08] explained "how probabilistic models (Bayesian networks) can be built and used to evaluate quality characteristics. The structure of the networks is defined by refinement of existing models, where the parameters (probabilities and probability tables) are set using expert judgment and fuzzy clustering of empirical data". **Vaucher et al.** [VBSH09] presented "a general approach to recommend improvements to Web applications. The approach uses a meta-heuristic algorithm to find best sequence of changes given a quality model (proposed by Haydar et al.) responsible to evaluate the fitness of candidate sequences". **Malak et al.** [MSBB10] presented "a probabilistic approach for building Web quality models and the associated assessment method. The proposed approach is based on Bayesian Networks". They illustrated the approach feasibility with "the important quality characteristic of navigability design". **Vaucher and Sahraoui** [VS10] proposed "an evaluation approach that

[26]Inspired by Bush's memex, these trails provide a structure to the returned results and provide users with contextual information not provided by traditional search facilities.

combines evaluations at the page level with the one of the web site by means of a page-importance weighing model". They illustrated their approach "with the particular characteristic of navigability".

NB: In the following chapter, we will explain in detail our Multi-level Model that aims to assess website navigability. Following the studies of Malak and Vaucher, we also use a probabilistic approach based on Bayesian Networks.

Chapter 3

Multi-level Model and Related Tool

In this chapter, we present the quality model of our study that aims to assess website navigability. We call it the *Multi-level Model*. It extends previous work of Malak [MSBB10] and Vaucher [VS10].
We first define the problem we want to model and justify why we study a "Quality Model" to assess website navigability. We use a probabilistic approach and discuss how Bayesian Belief Networks (BBNs) are used.
Second, we briefly present the reference navigability model of Malak. Then, we detail our *Multi-level Model*. We explain the three "cascading models" used to describe different levels of user navigation (a "Page-level Model" that computes a navigability score for a webpage, a "Composition Model" that produces an aggregate value of quality for all pages, and a "Site-level Model" that combines the assessment of site-level navigation elements with the composition model output).
We precisely contribute to extending the "Composition Model". We detail what we have added with the intention of completing this model.
Finally, we present how we have implemented the *Multi-level Model* (in the form of a Java program).

3.1 Problem Statement

In order to assess website navigability, we first define what we want to model. Graph theory notations are used to describe the navigability issue. We further use these notations to implement our *Multi-level Model*.

We explain why we use a Quality Model to assess website navigability. Our *Multi-level Model* is based on a probabilistic approach. We explain this choice and how we implement it by means of a Bayesian Belief Network (BBN).

3.1.1 Graph Theory Notations

In this work, we define website navigability as "the ease with which users can locate and access relevant information" [VS10].

We can use graph theory notations to do some thinking about the navigability issue. In order to assess its navigability, a website is viewed as a directed graph. We use the following notations:

- $G \Rightarrow \langle V, E \rangle$: the directed graph representing the website;
- V: the set of vertices representing the pages;
- E: the set of directed edges representing links between pages.
- (u, v): an edge that represents a link from the page u to the page v. Vertex u, is called the head of the link and v, the tail.
- For vertex u, the *out-links* is the set of links with u as the head, representing the links to the other pages.
- For vertex u, the *in-links* is the set with u as the tail, representing the links to u from the other pages.

A user requiring information located at page p_{dest} needs to find a path $(p_1, p_2, ..., p_{dest})$ in G that takes him from his origin p_1 to his destination p_{dest}. In terms of the graph, this is a greedy path-finding problem where at any given page a user needs to figure out which out-link leads him closer to his destination [VS10]. We use these notations to implement our *Multi-level Model*. However, we do not tackle our problem as a pure "Graph Theory" or "Fundamental Programming" issue. Our *Multi-level Model* is based on practical ways of navigating a website. We now explain what we need to consider in order to build a navigability model.

3.1.2 What to model

A user typically has two options to be successful in doing a navigation task.

First, he can go from page to page by following links in order to explore the website. A site with good navigability should ensure that few steps are required to reach any destination. Some potential navigation difficulties arise due to pages with inadequate link identification (*e.g.* bad anchor text) or to pages that overwhelm the user with too much information (*e.g.* he needs to scroll down to find the correct link).

Second, he can directly access a certain webpage using other navigation elements, such as a search engine or a site map. By using the search engine, the user jumps directly to a new page. Exploration is useful even if the site has been indexed by a search engine because, lacking the adequate keywords, a user may not find the page he needs. Both methods of navigation are thus complementary.

Consequently, the general navigability assessment of a website needs to take into consideration both ways of navigating the site. We need to modelize two navigation processes:

- **a page-level navigation process** that consists of searching a piece of information in a webpage or the good link to follow from the webpage to another one.
  The model should assess how easy it is to find a piece of information in a webpage or the appropriate link to follow on this page. The model should combine this to the probability that a user will be on that page.
  We divide this page-level navigation process into two separate models: a page-level model and a composition model (both models are explained in detail in the third section of this chapter).
- **a site-level navigation process** that consists of using global navigation elements that allow the user to jump directly to a new page.
  We propose a site-level model (explained in detail in the third section of this chapter) to take into account this site-level navigation process.

Before we explain in detail our *Multi-level Model*, we justify why we use a "Quality Model" to assess website navigability. We then describe why we resort to a probabilistic approach and use Bayesian Belief Networks.

3.1.3 Quality Model Approach

As previously introduced (cf. *"Introduction"* and *"State of the Art"*), a Quality Model is "a defined set of characteristics, and of relationships between them, which provides a framework for specifying quality requirements and evaluating quality" (ISO/IEC 25000:2005 definition). Web Quality Models are Quality Models that naturally focus on web quality.

We saw in the second chapter that "web quality" and "usability" assessment is traditionally done thanks to a survey of users who assess how easy it is to perfom specified navigation tasks. De Marsico and Levialdi [ML04] remind us that "user questionnaires are one of the most typical and consolidated tools to evaluate user interfaces" and they summarize some of the most popular sources of usability questionnaires.
Quality models are used as an alternative to surveys that are very expensive and cannot be fully integrated into a continuous development process. Ivory and Hearst [IH01] explain that the "automation of usability evaluation has several potential advantages over nonautomated evaluation, such as reducing the cost of usability evaluation or incorporating evaluation within the design phase of user interface development". These models have to assess how a user would react to a website. Web Quality Models can be built using results of a literature review or using data from surveys.

Malak [MSBB10] adds that "Usually software (including the Web) quality models are defined in terms of quality characteristics (and subcharacteristics) organized hierarchically. These models are generally used to assess software artifacts by measuring factors and by recursively using weights at different levels to evaluate each characteristic/subcharacteristic". Malak notices that these "strict hierarchical models" pose several problems in the context of Web applications: it is difficult to "define widely applicable threshold values, to give a meaning to weights when combining aspects of different nature, to represent a quality criterion that impacts more than one quality characteristic, and to provide adaptation/calibration mechanisms". Therefore, she uses a probabilistic approach and Bayesian Belief Networks to attenuate these problems.

3.1.4 Probabilistic Approach

Assessing navigability is a highly uncertain and subjective process. There are different types of users and their assessment depends on their experiences and opinions on how a site should be organised. We consequently need an approach capable of modeling uncertainty.

Malak proposed a probabilistic approach for building Web Quality Models. She precisely used Bayesian Belief Networks (BBNs) to manage the problems related to "strict hierarchical models" (cf. *Quality Model Approach*, just above). Malak [MSBB10] noticed the following BBNs advantages:

- "There is no need for defining satisfaction levels and threshold. Indeed, the use of probabilities allows dealing with metric values in a continuous way. Moreover, the influence of different criteria is defined as conditional probabilities that could be learned rather than weights for a linear composition.
- Graphical models are an intuitive visual representation for interdependent criteria. It is possible to represent multiple relationships, that is, a criterion that may impact many criteria.
- Bayesian Networks are known as good tools for solving prediction problems that involve reasoning with uncertainty.
- Bayesian Networks can be customized to take into account a particular context by modifying the parameters (probabilities)".

Probabilistic modeling is based on Bayes' conditional probability theorem. This theorem combines the inherent probability of an output (A) with its dependence on inputs (B) as well as the probability of (B) occuring. This is expressed by the following equation:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{3.1}$$

where $P(\alpha|\beta)$ denotes the conditional probability of α given β.
"Bayes' rule enables to do probability calculations on chains of events that influence each others probability of occurring. This is exploited in the Bayesian Belief Network (BBN)" [Huy02].

Following Malak, we use Bayesian Belief Networks (BBNs) to implement a probabilistic approach. First, we briefly present BBNs: definition, notations, and reasons to use them. Then, we introduce the reference navigability model of Malak before giving a complete description of our *Multi-level Model*.

3.1.5 Bayesian Belief Networks

A Bayesian Belief Network is a probabilistic graphical model that represents a set of random variables (nodes) and their conditional dependencies (edges) via a directed acyclic graph (DAG).
“A random variable denotes an attribute, feature, or hypothesis about which we are uncertain. Each random variable has a set of mutually exclusive and collectively exhaustive possible values. That is, exactly one of the possible values is or will be the actual value, and we are uncertain about which one it is.
The graph represents direct qualitative dependence relationships; the local distributions represent quantitative information about the strength of those dependencies. The graph and the local distributions together represent a joint distribution over the random variables denoted by the nodes of the graph.
One of the most important features of Bayesian Networks is the fact that they provide an elegant mathematical structure for modeling complicated relationships among random variables while keeping a relatively simple visualization of these relationships”[1]. BBNs do not only allow you to visualize but also to do some thinking.

BBNs Notations

A Bayesian Belief Network is a directed acyclic graph (DAG). Each network node represents a variable. Edges represent conditional dependencies.

$X = X_1, X_2, \ldots, X_n$: The set of random variables (n = number of variables).

Vertices: Each vertex corresponds to a random variable. It is an observable (and measurable) concept or a decision point, given inputs defined by parents. With discrete variables, each variable X_i is configured with a Conditional Probability Table (CPT) that establishes its probability distribution given the values of its parent nodes.

Edges: The edges represent causal relations between vertices. Each edge connecting two vertices indicates a probabilistic dependency of the head, called parent, on the tail, called child. (The edges allow a developer to interprete the results of an evaluation: the output is caused by this input).

[1]From `http://www.pr-owl.org/basics/bn.php`; (Date of access 2011-07-18).

The joint distribution of X can be written as[2]:

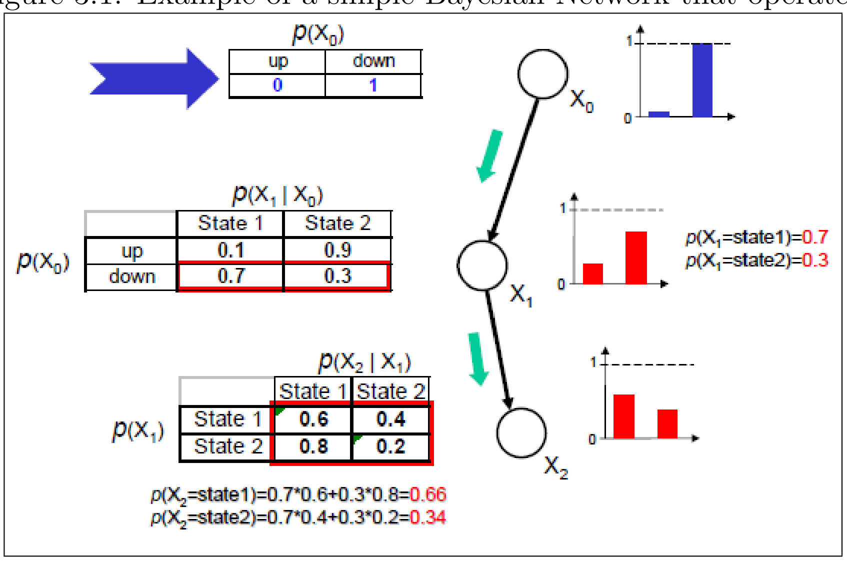
$$P(X_1, ...X_n) = \prod_{1..n} P\big(X_i \,\big|\, parent(X_i)\big) \tag{3.2}$$

where *parent*(X_i) is the set of parents of X_i (i.e. those vertices pointing directly to X_i via a single edge).

The random variables represented in the graph are only conditionally dependent on their parents. This structuring can be done either automatically using heuristics found in the literature or manually to correspond to a specific decision process. In our case, we keep the CPTs used by Malak in the reference navigability model [Mal07].

Figure 3.1 shows how a simple Bayesian Network operates[3]. In the example,

Figure 3.1: Example of a simple Bayesian Network that operates.



we know the state of the variable X_0 (*down*). We calculate the probabilities for the variable X_1 (according to its parents: X_0). Knowing that $p(X_0 = down) = 1$, the CPT of X_1 gives $p(X_1 = State1) = 0.7$ and $p(X_1 = State2) = 0.3$. Similarly, the probabilities for the variable X_2 are calculated according to its parents: X_1.

[2]In the study of probability, given two random variables Y and Z defined on the same probability space, the joint distribution for Y and Z defines the probability of events defined in terms of both Y and Z.

[3]According to Philippe Weber, *Application de la modélisation par Réseaux Bayésiens à la sûreté de fonctionnement*, Nancy Université, 2008.

Reasons to use BBNs

Bayesian Belief Networks (BBNs) are frequently used to model quality. BBNs have two major advantages. First, BBNs can model the uncertainty inherent to decision-making by using Bayes' theorem. Second, the model can guide improvements because the evaluation process can be used backwards: for a desired output, it can identify the required inputs.

BBNs are also better than other models (e.g. artificial neural network, expert system, decision tree, data analysis model (linear regression), fault tree analysis or logic models) thanks to some particular aspects such as: [NWL+04]

- Knowledge acquisition: It is possible to gather and merge knowledges of different kinds in the same model. These different knowledges can come from experience feedback, assessment (logical rules, equations, statistics, etc.), observations.
- Knowledge representation: The graphical representation of a Bayesian Network is explicit, intuitive, and understandable by non-specialists. It is easier for the model validation, possible evolutions and, above all, for practical use.
- Knowledge use: A Bayesian Network is multi-purpose: with the same model, it is possible to assess, plan, diagnose, or optimize decisions.

Most of all, available softwares that deal with BNBs have a good quality level. These tools have more or less evolved features: probabilities and network structure apprenticeship, possibility to integrate continuous, utility, or decision variables, etc.

Now, we introduce the reference navigability model of Malak (based on a probabilistic approach, implemented by means of a Bayesian Network). After that, we explain in detail our *Multi-level Model*.

3.2 Reference Navigability Model

Our reference model is the Bayesian Belief Network presented by **Malak et al.** [MSBB10]. We call it the "Reference Navigability Model". **Malak et al.** used a Goal - Question - Metric (GQM) Approach[4] in order to refine navigability characteristics collected from ten sources (models, standards, guidelines, and best practices).

3.2.1 Brief Description

In this model (illustrated in Figure 3.2), the fact that a page is easily navigable directly depends on three sub-characteristics (the *Navigability* node "refers to the presence in this page of specific elements that allow the user to identify and locate the information required and to access rapidly this information through the use of hyperlinks" [MSBB10]):

- The *Locate* node subnetwork describes the ease to locate information.
- The *Down. Speed* node is the time required to download a page. In a second version of the model, Malak replaced the "download time" by the "page size".
- The *Bind* node subnetwork describes the facility to access the information on the destination page.

Each intermediate node directly depends on its sub-nodes. The leaf nodes of the Bayesian Network are input metrics (measure or binary value). "The quality assessment of a Web page is done in three steps: (1) measurement, (2) derivation of input probabilities, and (3) Bayesian Network evaluation and improvement scenario exploration"[MSBB10]. We do not explain this model in detail. Additional information can be found in [MSBB10].

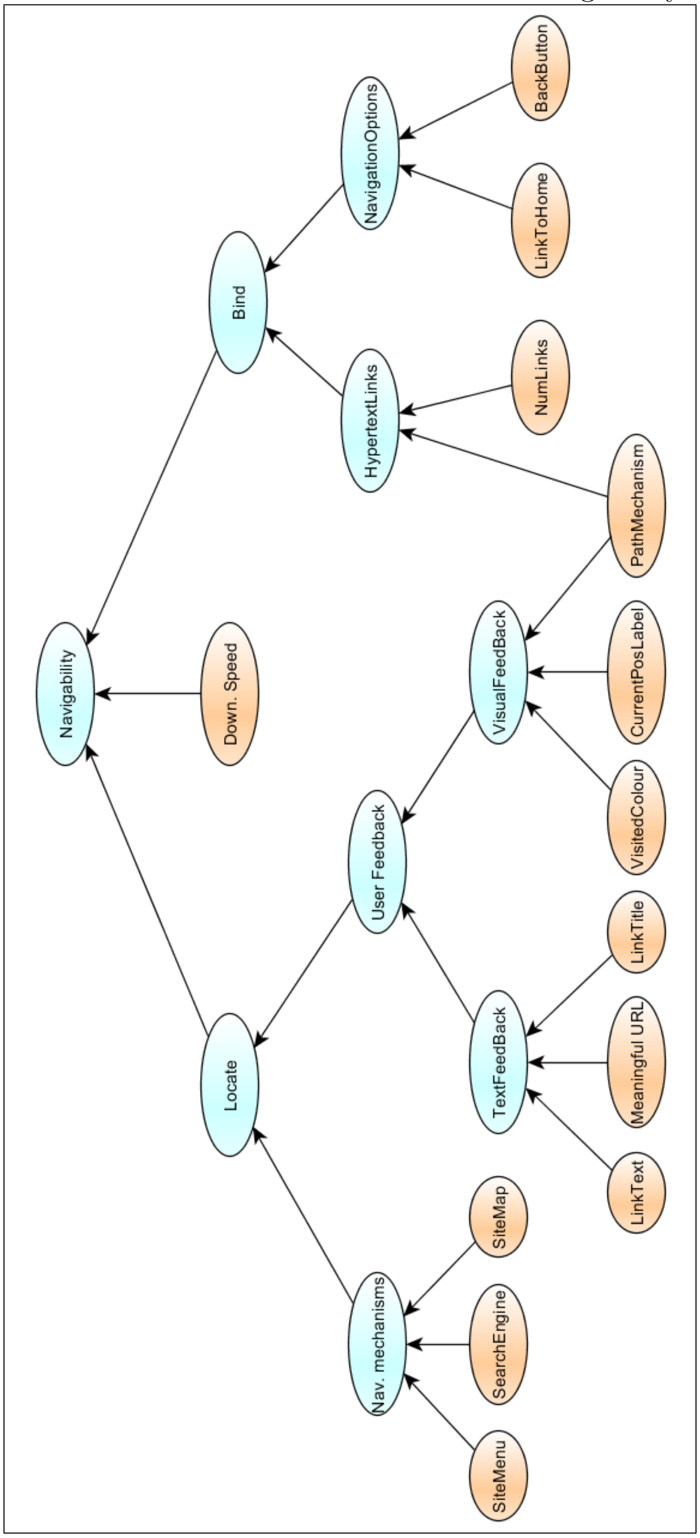
In the following section, however, we give a complete description of our "Page-level Model" (directly inspired from this "Reference Navigability Model").

3.2.2 Limits

The "Reference Navigability Model" is a page-level model. It aims to assess a webpage. This model uses several metrics describing a webpage and a website.

[4]GQM paradigm was defined by **Basili et al.** [BCR94]. It helps to review and verify the criteria classification, then to define and assign metrics to different criteria and subcriteria.

Figure 3.2: BBN structure of the "Reference Navigability Model".



Vaucher et al. [VS10] criticized the presence of website metrics at this level. They chose to separate webpage assessment from website assessment. We keep their multi-level approach.

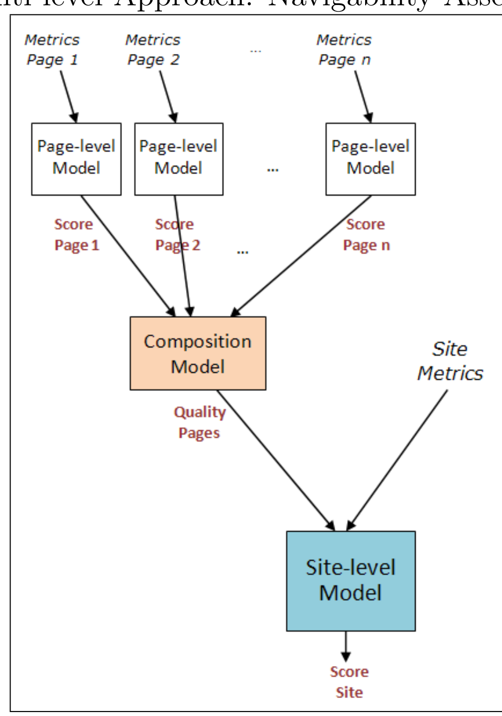
In the "Reference Navigability Model", site-level elements are metrics for the *Nav. mechanisms* node. In our *"Multi-level Model"*, we exlude these elements from the page-level model. Consequently, we combine the page-level model with a site-level model that takes into account the site-level navigation elements. In the next section, we explain in detail our multi-level approach.

3.3 Multi-level Approach

In order to take into account a multi-level approach, we develop three distinct quality models: a page-level, a composition, and a site-level model. Our *"Multi-level Model"* combines these three "cascading models" to assess website navigability.

The navigability assessment process that we define is illustrated in Figure 3.3. Now, we explain in detail these models and how they are combined together.

Figure 3.3: Multi-level Approach: Navigability Assessment Process.



3.3.1 Page-level Model

The Page-level Model aims to assess the navigability of a webpage. It is directly derived from the "Reference Navigability Model", with some modifications. We eliminate five nodes: 3 metric nodes (SiteMenu, SearchEngine, SiteMap) and 2 decision nodes (Nav. mechanisms, Locate).

Compared with the "Reference Navigability Model", the subgraph corresponding to (site-level) navigation mechanisms (*Nav. mechanisms* node and its 3 sub-nodes: *SiteMenu; SearchEngine; SiteMap*) is moved directly to the "Site-level Model". Consequently, we eliminate the *Locate* node. Indeed, the purpose of an intermediate node is to simplify computation and this node becomes thus unnecessary. In the *Navigability* CPT, the *Locate* node presence is directly replaced by the results of the *UserFeedback* node. Figure 3.4 presents the Bayesian Network structure of the Page-level Model.

Page-level Model Presentation

The Page-level Model is represented by a BBN, that we can see as a quality tree. The root node (*Navigability* node) is computed on the basis of three sub-nodes, which are navigability sub-characteristics. Then, the fact that a page is navigable directly depends on three sub-characteristics:

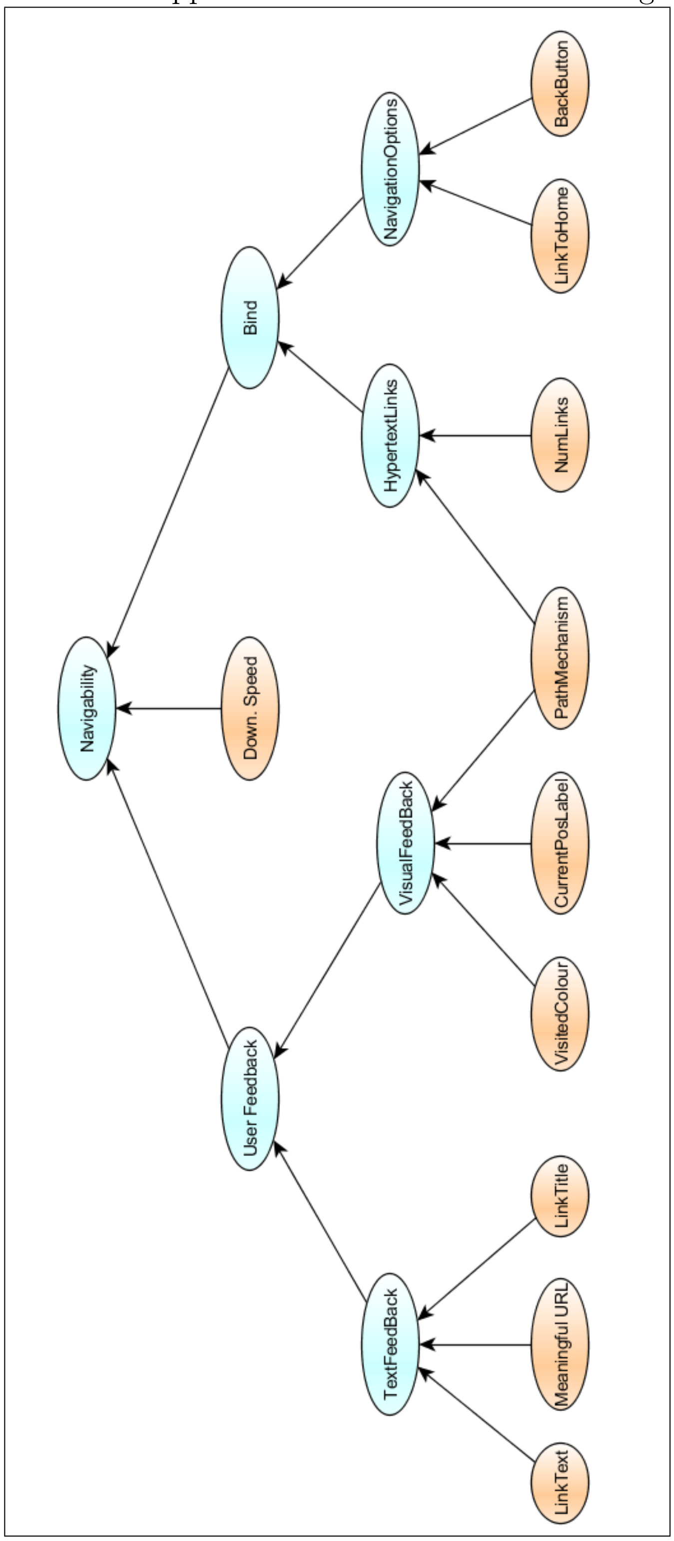
- UserFeedback: the ability of a user to identify the correct link to follow ;
- Bind: the user access to available navigation mechanisms ;
- Down.Speed: the size of a downloaded page.

Intermediate decision nodes as *UserFeedBack* and *Bind* nodes depend on other sub-characteristics (for example, *TextFeedback* and *VisualFeedback* are sub-characteristics of the *UserFeedback* node). The tree leaves are the input nodes and consist of page-level navigability metrics. Now, we describe all the nodes of our *Page-level Model*.

Metric Nodes

Table 3.1 presents the 10 navigability metrics (input nodes) of the Page-level Model. "Transforming numbers into probabilities can be done in different ways depending on the nature of the criterion" [MSBB10].

Figure 3.4: Multi-level Approach: BBN structure of the Page-level Model.



In the case of binary criteria, the variable is measured by its presence or not on the page (two states: "Yes" et "No"). For example, if a *PathMechanism* is available on the webpage, we assign a probability of 99% to the state "Yes" and 1% to the state "No" (the probabilities should not be a zero in a Bayesian Network to maintain the uncertainty).
In the case of criteria with measurable levels of existence (e.g. *LinkTitle* node), the probability of each state is determined by a percentage. For example, if 30% of links have titles, in a given application, we assign a probability of 0.3 to state "With title" and 0.7 to the state "Untitled".
Certain input nodes (e.g. *NumLinks* node) are criteria with infinite measurable values. They are transformed into probabilities by means of fuzzy logic and an approximation method. This is explained in detail in [MSBB10].

Figure 3.5: Page-level Model: VisitedColour node (example where the colour of a visited link has changed).

Intermediate Nodes

In our Page-level Model, 7 nodes are "composed" by other sub-nodes. The root node is the *Navigability* one. It is preceded by 6 intermediate nodes. We describe these composite nodes below. Their Conditional Probability Tables (CPTs) are presented in Appendix A, section Page-level Model: CPTs of the BBN.

TextFeedBack node: This intermediate node computes the probability that a user has a good feedback thanks to the text quality of the links and of the webpage URL. The CPT takes into account 3 metric nodes as inputs: *MeaningfulURL*, *LinkTitle* and *LinkText*.

Table 3.1: Input nodes of the Page-level Model.

Input node (metric)	**Definition**	**Measure**
Down. Speed	The time required to download a page, as measured by the size of the page.	Webpage size
LinkTitle	Ratio of links with titles.	$\frac{\textit{Links with titles}}{\textit{All the links}} * 100$
LinkText	Ratio of links with (significant) text.	$\frac{\textit{Representative links}}{\textit{All the links}} * 100$ *Representative links = links with significant text - links with the form "next", "click here", "more", "previous", etc.*
MeaningfulURL	Significance of page URL.	Is the page URL relative to the page or not ? (binary measure)
VisitedColour	Visited links change colour *(Figure 3.5 gives an example from `http://www.allocine.fr`).*	Is the colour of the visited links different from the others? (binary measure)
CurrentPosLabel	Indication of location on the website *(Figure 3.6 gives an example from `http://edition.cnn.com`).*	On the page, is there an indication of location? (binary measure)
PathMechanism	Presence of breadcrumbs *(Figure 3.7 gives an example from `http://www.futureshop.ca`).*	On the page, is there a path indicator or not, often at the level of the menu/navigation bar ? (binary measure).
NumLinks	Number of links on the page.	Count: whole number.
LinkToHome	Link to the home page.	On the page, is there a link to home or not ? (binary measure)
BackButton	Support for the Back Button.	On the page, buttons "next" and "previous" are still active or not ? (binary measure)

VisualFeedBack node: This intermediate node assesses the probability that a user can easily know where he is on the website (by means of "visual" navigation mechanisms). The CPT takes into account 3 metric nodes: *CurrentPosLabel*, *PathMechanism* ("breadcrumbs") and *VisitedColour*.

Figure 3.6: Page-level Model: CurrentPosLabel node (example of a webpage with CurrentPosLabel).

Figure 3.7: Page-level Model: PathMechanism node (example of a webpage with breadcrumbs (PathMechanism).

UserFeedBack node: Based on these two previous intermediate nodes, the *UserFeedBack* node is defined as the probability that a user has a good textual and visual feedback. The CPT has 2 intermediate nodes as inputs: *VisualFeedBack* and *TextFeedBack*.

HypertextLinks node: This intermediate node computes the probability that the webpage has a good quality at the level of hypertext links. The CPT takes into account 2 metric nodes as inputs: *PathMechanism* and *NumLinks*.

NavigationOptions node: This intermediate node assesses the probability that the webpage presents good navigation options, in terms of a link to the home page and back button availability. The CPT takes 2 metric nodes as inputs: *LinkToHome* and *BackButton*.

Bind node: This intermediate node assesses the probability that a user can easily bind the current webpage with the rest of the whole website. The CPT takes into account 2 intermediate nodes as inputs: *NavigationOptions* and *HyperTextLinks*.

Navigability node: The Bayesian Network root node computes the probability that the webpage has a good navigability level. The CPT takes into account the *Bind*, the *DownSpeed*, and the *UserFeedBack* nodes.

As our Page-level Model is directly derived from the "Reference Navigability Model" of Malak, more information can be found in [MSBB10].

Example of a Webpage Navigability Assessment

We show how the BBN works in order to assess a webpage. We compute here the input nodes "manually". We use a Baysian Belief Network editor (Hugin[5]) to infer the network. For example, we want to assess the following webpage: `http://www.cedric-klapisch.com/films/poupeesrusses.html` (cf. Figure 3.8).

Input nodes :

- LinkText = 1 (Ratio of links with texts = 12/12 = 1).
- MeaningfulURL = 1
- LinkTitle = 1
- VisitedColour = 0
- CurrentPosLabel = 0
- PathMechanism = 0

[5]Hugin Expert A/S (`http://www.hugin.com/`) is a provider of software for advanced decision support based on Bayesian Networks.

Figure 3.8: Example of a webpage that we want to assess.

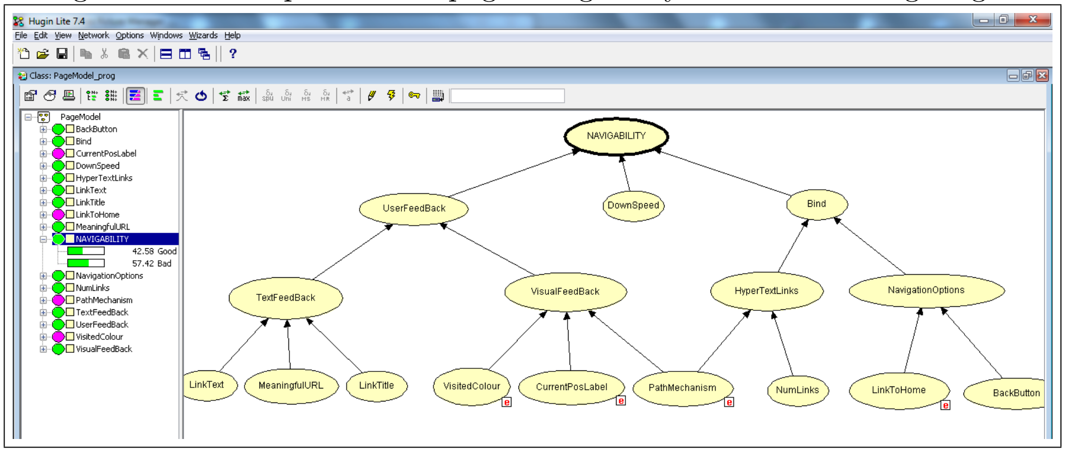
- NumLinks = 12
- LinkToHome = 0
- BackButton = 1

We illustrate with the Hugin tool how the navigability score is computed for this webpage (cf. Figure 3.9). We observe that the score of the *"VisualFeedBack"* intermediate node is very bad because of the lack of its subnodes (no path mechanism, no colour change of the visited links, and no current position label). Consequently, the navigability score for the page is quite low (42.58%).

3.3.2 Composition Model

This model aims to assess the navigability for all the pages. It computes the importance of each page and combines it with each webpage navigability score. The *page importance* notion is introduced to model the probability that a user will transit by that page to reach the desired page. (We remind the navigability

Figure 3.9: Example of a webpage navigability assessment using Hugin.



definition: "The navigability of a website is a measure of how easily a user can locate and access the information he needs").

In previous works, Vaucher et al. [VS10] presented an algorithm based on random walks to compute page importance. We complete this approach by selecting different algorithms that compute page weights on the basis of various strategies. The experiment will allow us to analyse the impact of such algorithms (cf. next Chapters).

Now, we explain the different weighting algorithms used to complete the Composition Model.

Weighting Algorithms

Many existing graph-based algorithms assess the importance of a node given the topology of the graph. We remind that a website can be viewed as a directed graph to assess its navigability: each node represents a webpage and edges represent links between webpages (cf. *"Problem Statement"* section; *Graph Theory Notations*).

We present the "Visit Probability" algorithm. It was the only strategy to aggregate webpages scores in the first version of the Multi-level Model presented by Vaucher [VS10]. In order to complete our Composition Model, we select some other weighting algorithms, based on [WS03]. We also add a "Simple Mean" strategy that computes the mean navigability score for all webpages, without weighting algorithm.

Visit Probability: This algorithm is based on random walks in order to compute the importance of each webpage. A user starts from the home page. Then, for each given page p, it assumes that a user will follow a randomly chosen link with a uniform probability ($1/outlinks(p)$). The algorithm is based on a breadth-first search and is presented in Algorithm 1.

```
Inputs  : home: the start page
Inputs  : outlinks: a vector of outlink for a page
Outputs: weight: a vector describing the relative weight of a page
Q ⇐ empty − queue
mark[home] ⇐ visited
weight[home_page] ⇐ 1
Clicks ⇐ 0
enqueue home into Q
repeat
    dequeue page from Q
    foreach Link(page,v) ∈ outlinks(page) do
        weight[v] ⇐ weight[v] + weight[page]/|outlinks|
        if mark[v] ≠ visited then
            mark[v] ⇐ visited
            enqueue v into Q
        end
        Clicks ⇐ Clicks + 1
    end
until Q is not empty;
forall the v ∈ weight do
    v ⇐ v/Clicks
end

return Visits
```

Algorithm 1: Visit probability

Simple Mean: This strategy aggregates webpages navigability scores by means of a simple arithmetic mean. It means that we give the same importance to all the webpages.

$$SimpleMean = \frac{\sum_{i=1}^{n} navScorePage_i}{n} \tag{3.3}$$

where $navScorePage_i$ is the navigability score of the i^{th} webpage; and n is the number of webpages.

Betweenness: Betweenness is a centrality measure of a vertex within a graph. Vertices that are on many *shortest paths* between other vertices have higher betweenness than those that are not. "High centrality scores thus indicate that a vertex can reach others on relatively short paths, or that a vertex lies on considerable fractions of shortest paths connecting others" [Bra01]. BC(v): "The betweenness centrality of a vertex $v \in V$ is the sum over all pairs of vertices $s, t \in V$, of the fraction of shortest paths between s and t

that pass through v" [NMVN05]:

$$BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3.4}$$

where σ_{st} is the number of shortest paths from s to t ;
and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through v.
[Bra01] and [NMVN05] present algorithms to compute the vertex betweenness centrality measures of all the vertices in a graph.

Random Walk Betweenness: Computes betweenness centrality for each vertex in the graph. "The betweenness values in this case are based on random walks, measuring the expected number of times a node is traversed by a random walk averaged over all pairs of nodes"[6].

Weighted NI Paths: This algorithm measures the importance of nodes "based upon both the number and length of disjoint paths[7] that lead from the root node to a given node. It uses heuristic breadth-first search to find the node-disjoint paths between two nodes"[8].
[WS03] explains in detail how to compute relative importance using weighted paths.

MarkovCentrality: "The idea is to start at some random position on the graph and then take an infinite-length random walk on the graph. By random walk, it means that the walker starts at some node and then randomly chooses an outgoing edge to follow to the next node. The process then repeats itself. The Markov Centrality of a node is the probability that, at the end of such an infinite process, the walker will find itself on a particular node" [BMB10].
"Nodes that are more 'central' in a network, i.e., closer to the center of mass, have higher ranking than those that are less central" [WS03] (that explains in detail the Markov Centrality).

[6] According to `http://jung.sourceforge.net/doc/api/edu/uci/ics/jung/algorithms/importance/RandomWalkBetweenness.html`

[7] Node-disjoint paths are paths that have neither edges or nodes in common, i.e., no node or edge can be used more than once [WS03].

[8] According to `http://jung.sourceforge.net/doc/api/edu/uci/ics/jung/algorithms/importance/WeightedNIPaths.html`

K Step Markov[9]: This strategy computes the importance of each node by means of fixed-length random walks starting from the root node. It computes the stationary probability of being at each node. Specifically, it computes the relative probability that the markov chain will spend at any particular node, given that it starts at the root node and ends after k steps. [WS03] explains in detail the k-step Markov Approach.

Page Rank: Assigns scores to each vertex according to the Page Rank algorithm that was initially described in [PBMW99]. It is used by Google.
In our case the Page Rank algorithm has to assign a weight to each page of a website. Each page receives a score proportional to the number of times that a user goes through this page, when he surfs all over the website and randomly clicks on links appearing on each webpage.
We use the implementation provided by the Jung Framework[10]. “The score for a given vertex may be thought of as the fraction of time spent ’visiting’ that vertex (measured over all time) in a random walk over the vertices (following outgoing edges from each vertex). PageRank modifies this random walk by adding to the model a probability of jumping to any vertex”[11].

Hub (HITS): The HITS (Hyperlink-Induced Topic Search) algorithm is “based on the relationship between a set of relevant authoritative pages and the set of ’hub pages’ that join them together in the link structure” [Kle99]. “A good hub page is one that points to many good authorities; a good authority page is one that is pointed to by many good hub pages[12]”.
We choose to compute the hub value of each page (that assesses the value of its links to other pages) rather than the authority value (which focuses on the content of the page) in order to match at best with our navigability definition. We use the implementation provided by the Jung Framework[13].

[10] From `http://jung.sourceforge.net`

[11] From `http://jung.sourceforge.net/doc/api/edu/uci/ics/jung/algorithms/scoring/PageRank.html`

[12] According to `http://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html`

[13] From `http://jung.sourceforge.net/doc/api/edu/uci/ics/jung/algorithms/scoring/HITS.html`

Algorithms Integration in the Composition Model

The contribution of a page to the navigability of the site is determined by the function *weighted_importance* (Equation 3.5)

$$weighted\_importance(page) = proba(page) * nav(page) \tag{3.5}$$

where $proba(page)$ is the probability that a user will visit $page$ (the page weight) and $nav(page)$ is its navigability score as assessed by the Page-level Model.

The total page navigation score $ComposedPageNav$ of a site is calculated as the average *weighted_importance* considering all the pages of the site (Equation 3.6). It is the "Composition Model" output.

$$ComposedPageNav(site) = \frac{\sum_{p \in site} weighted\_importance(p)}{|site|} \tag{3.6}$$

3.3.3 Site-level Model

The Site-level Model is divided in two parts. The *Site Navigability* node is computed by means of a CPT that combines the information from the site and from the page-level assessments. There are thus two different input nodes for this CPT:

The Nav. Mechanisms node is the result of a subgraph describing site-level mechanisms (assessment of the navigation menu; the intern search function and the site map);

The ComposedPageNav node corresponds to the output of the Composition Model. It gives an aggregated navigability score for all the webpages (given that each webpage has its own weight, according to a selected importance function).

The CPT returns P(*SiteNavigability* = Good) = 99% if both its inputs, *ComposedPageNav* and *Nav.Mechanisms*, are "good" (*ComposedPageNav* = good and *Nav.Mechanisms* = good). If either one of the inputs is good, P(*SiteNavigability* = good) = 70%. The Conditional Probability Tables (CPTs) of the site-level model are presented in Appendix A, section Site-level Model: CPTs of the BBN.

3.4 Related Tool

In this section, we present our Java program that implements the theoretical model. Initially, this program was developed by Stéphane Vaucher. Then, we added some features.
Our tool aims to assess website navigability as described in the *Multi-level Model*. We briefly explain the program structure and how it works.

3.4.1 General Explanation

The program takes the form of a webcrawler that aims to assess the navigability of a website. First, the program receives a list of URLs as input. The program has to download a given number of pages for each website. For each website, it works as follows. When the program has the pages, it runs the assessment process. Starting at the homepage, the program computes the navigability metrics for each webpage (according to the page-level model). The importance of each page is also computed thanks to different algorithms (according to the composition model). The website score is computed according to the site-level model. Finally, the program displays and records the navigability score of each webpage and the navigability score of the website.

3.4.2 Program structure

We summarize the program structure and its division into packages. There are three main packages called:

- **metrics:** This package contains 4 sub-packages:
  - *web*, that contains classes used to determine if the website or the webpage contains a certain navigation element or not.
  - *bayes*, that contains classes that
    - Model the transformation of a binary value into a distribution node;
    - tranform numeric values to a membership set;
    - take discrete fuzzy metrics as input and return a fuzzy output and compute final score for a web page or for a website;

    NB: Bayesian Network structures are represented by means of XML files.
  - *core*, that contains classes that manage the computation of probability distribution.

  - *statistics*, that contains classes that manage the statistical thresholding and descriptive statistics.

- **quality:** This package manages different metric nodes and the computation of a Bayesian Node.
- **webscrape:** This package contains the main class (*WebCrawler*) used to download a website and to launch the navigability assessment. Moreover, this package manages the choice and the implementation of the weighting algorithms. Finally, it contains classes that represent a webpage, a website, and the website graph.

3.4.3 Implementation Choices

We explain our choices about the programming language and the main frameworks we used.

Programming Language: We use Java to develop our tool. This choice allows us to easily integrate different frameworks such as HTMLUnit, Jung and BNJ.

Frameworks: The main frameworks used to develop our tool are:

- **HTMLUnit** We use HTMLUnit to download websites and to enable us to detect navigation elements.
  HtmlUnit[14] is a "GUI-Less browser for Java programs that models HTML documents and provides an API that allows to invoke pages, fill out forms, click links, etc. It has fairly good JavaScript support and is able to work even with quite complex AJAX libraries".
  We choose to base our crawler on HtmlUnit because it includes a JavaScript interpretor, and is known to support many JavaScript pages. This is important for modern sites (like those that are AJAX-based) as it can handle JavaScript events that are executed when a button/link is clicked. It can therefore download the content of most (non-Flash) websites.
- **Jung (The Java Universal Network/Graph Framework)** We use Jung to implement certain weighting algorithms according to our composition model.
  Jung[15] is "a software library that provides a common and extendible language for the modeling, analysis, and visualization of data that can be

[14] According to `http://htmlunit.sourceforge.net/`; (Date of access 2011-02-12).
[15] According to `http://jung.sourceforge.net/`; (Date of access 2011-02-12).

represented as a graph or network. It is written in Java, which allows JUNG-based applications to make use of the extensive built-in capabilities of the Java API, as well as those of other existing third-party Java libraries".

- **BNJ (Bayesian Network tools in Java)** We use BNJ in order to work with Bayesian Networks.
  BNJ[16] is "an open-source suite of software tools for research and development using graphical models of probability".

3.4.4 Limits

The crawler can not support many server-side operations. The main one concerns the use of server-side scripting. On large sites, pages are built dynamically using templates. Our crawler can only see the result of this scripting. We do not try to support an analysis of this scripting.
Many sites redirect requests from one page to another with a different URL (using HTTP redirections, JavaScript, or an HTML meta refresh tag). The result of this is that many pages with the different URLs can be redirected to the same page. HttpUnit supports these types of redirections, but to build our navigation model, we needed to use a coherent URL. We perform a limited canonicalisation of the URLs of the downloaded pages: we use an equivalence table of requested and redirected URLs.
Finally, flash websites are not supported by our tool.
During the experiment (cf. Chapters 4 & 5), we always check that our tool is able to assess the selected websites. Moreover, we check that the site-level navigation elements are correctly detected.

[16]According to `http://bnj.sourceforge.net/`; (Date of access 2011-02-12).

Chapter 4

Validity of the Multi-level Model: an Experiment

In the previous chapter, we described our quality model (the *Multi-level Model*) that aims to assess website navigability. We showed how we implemented the model into a tool (in the form of a Java program).

We now try to assess to what extent our model is valid. We conduct an experiment in order to compare human estimations about website navigability with navigability scores computed by our model.

In this chapter, we present the experiment conducted about website navigability assessment. First, we define this experiment according to the framework presented in [WRH$^+$00]: experiment definition (object of study, purpose, quality focus, perspective, context) and planning. The planning section focuses on the hypothesis formulation. We present the objectives and research questions we want to discuss. We also explain the variables selection, experiment design, and instrumentation. Then, we tackle the validity assessment of our experiment. Finally, we summarize the experiment operation. In the following chapters, we will describe and discuss the experiment results and findings.

4.1 Definition

We define the experiment via the "Goal Definition Template"[1], presented as:

Analyse "*Object(s) of study*"

for the purpose of "*Purpose*"

[1]The Goal Definition Template is based on the Goal-Question-Metric Approach [BCR94].

with respect to their "*Quality focus*"
from the point of view of the "*Perspective*"
in the context of "*Context*".

We are studying the Multi-level Model in order to validate its ability to simulate human judgments about website navigability. We can imagine the importance of such a model for a quality control manager in a business context. However, our "direct" point of view is the one of a researcher. We detail the definition of the experiment below.

4.1.1 Object of study

The studied entity of this experiment is a quality model (the *Multi-level Model*) that assesses the website navigability thanks to a multidimensionnal approach. This quality model is explained in depth in Chapter 3 (section Multi-level Approach); it is composed of three "cascading models":

A page-level model that computes a navigability score for a webpage (according to an assessment of the page-level navigation elements).

A composition model that computes an aggregated score for the global quality of all the webpages (according to the score of each webpage, from the *page-level model*, and a weighting strategy).

A site-level model that computes a navigability score for a website (according to the *composition model* result and an assessment of the site-level navigation elements).

In order to assess the validity of this model, we developed a Java program that implements the given quality model. The program is presented in Chapter 3 (section Related Tool).

4.1.2 Purpose

Our goal is to study the validity of the *Multi-level Model* defined in Chapter 3. The main question we want to answer is the following one: "Is our Multi-level Model able to assess website navigability?".

In concrete terms, we define two research objectives that we discuss in detail later:

- Evaluate whether or not the page-level model is able to accurately simulate human judgments.
- Evaluate whether or not the multi-level model can produce better estimates.

We explain our objectives and research questions in the *"Planning"* section.

4.1.3 Quality Focus

The primary effect under study in the experiment is the navigability of real websites. On the one hand, the quality of website navigability is computed according to the multi-level model (first, limited to the composition model; then, through the entire multi-level model). On the other hand, the quality of website navigability is assessed by a sample of web users (experiment subjects). In both cases, we focus on the quality of website navigability.

4.1.4 Perspective

Our "direct" viewpoint is the one of a researcher. We remind that we continue a previous work that was realised by Sofware Engineering researchers from the University of Montréal. Moreover, our navigability issue could be studied by a quality control manager who focuses on websites.

4.1.5 Context

In a business context, we can imagine a quality control manager who needs an effective navigability model in order to perform a quality audit (e.g. to test the quality of a new website, or a modification of the website, by means of a quality model to avoid the costs - resource and time - of a users' survey).
In our case, we try to improve and validate a quality model that aims to assess, as accurately as possible, the website navigability.

Dimensions of the context

We can add the four dimensions that characterize the experiment context.

- Off-line vs. On-line: **On-line**. Each subject has to carry out the experiment thanks to an online questionnaire.
- Students vs. Professional: **Students**. The subjects are for the greater part PhD researchers and Master students in Computer Science.
- Toy vs. Real problems: **Real problems**. The web navigability issue concerns all web users in their everyday surfing activities.

- Specific vs. General: **Specific**. Even if our study object takes place in a general concept (web quality), we focus on the navigability as a specific usability aspect. The experiment subjects belong to a specific group of users (in the Computer Science domain). We have to make this choice, even tough an ideal sample of subjects should represent all web surfers. But since such a group of subjects should be very large, it is not conceivable.

We start by explaining how the experiment concretely works. Then, we introduce the environment in which the experiment takes place and the experimental conditions. Finally, we define and describe our experiment subjects and objects.

4.1.6 How does it work?

First of all, we explain in a few words how the experiment works. A subject has to assess the navigability of a few websites. Each subject receives by e-mail the experiment explanation and a link to an online questionnaire. On this questionnaire the subject is invited to:

1. solve a navigation task on a given website. In concrete terms, the subject has to find a precise piece of information on the website (for example, an article or the price of a given item).

2. After his task, the subject has to fill in the questionnaire and say if he was successful or not. When he succeeds, he is invited to give the answer (or the webpage URL if the task was to find a complete article for example). We can thus make sure that the subjects have done the task seriously. When the subject did not find the answer, he can explain why he failed (too difficult task, website too difficult to use, absence of a certain navigation element, etc.).

3. Answer some questions on the questionnaire about his navigation session. The subject is invited to assess the usability of different navigation elements that were available (site-level elements, such as an internal search function and page-level elements, such as "link quality" for example). He has to assess the navigation elements on a 1 to 4 scale. He can of course indicate that a certain element was missing or not tested.

4. Write any additional comments he wants about the current task.

The same process is repeated for each navigability task with an identical questionnaire provided to the subject.
NB: Appendix B details the tasks performed by the subjects (section B.2).

4.1.7 Environment

We first conduct the experiment at the University of Montréal. In order to go further into our study, we lead a replication phase at the University of Namur. The replication phase should allow us to check if our first results and findings are confirmed or not. Except where otherwise specified, we assume the same setting for the first and replication phases of the experiment.

Experimental Conditions

The subject receives a document that he reads before starting the experiment. This document explains the goal and the rules of the experiment. It gives advice to well perform the experiment and provides definitions of some terms used in the questionnaire. Appendix B contains a copy of this document (section B.3).

The experiment is an online session. Each subject does the experiment on a classical workstation (desktop computer or laptop). We assume that the subject carries out the experiment on typical computer screens (*e.g.* 15 or 17 inches). The subject does not do the experiment on a smartphone or any other little screen. Moreover, we assume that subjects are sitting on a stool during the experiment. For each phase we organize a lab session that allows us to check if our subjects have a good understanding of the experiment goal or terms definitions.
Anyone who wants to take the experiment is free to stop it at any time. Each subject knows that the time needed to perform the experiment is not taken into account.

4.1.8 Subjects

For the first experiment phase, subjects are either members of the GEODES[2] team (researchers, Master's and PhD students) in the DIRO[3] of the Univerisity of Montréal, either Master's students in Computer Sciences from the University

[2]GEODES: The Software engineering group of the University of Montreal, `http://www.iro.umontreal.ca/~labgelo/main/index.php`.
[3]DIRO: Department of Computer Science and Operations Research, `http://www.iro.umontreal.ca/`.

of Montréal and Namur. Twenty-two subjects took part in the first experiment phase.
For the replication phase, subjects are Master students in Computer Sciences from the University of Namur. 24 subjects took part in the replication phase. None of them took part in the first experiment phase.

Language Knowledge

In order to distinguish potential groups between subjects at the language skills level, we first ask each subject to estimate his ability to use websites in English and in French (on a 1 to 5 scale).
Table 4.1 presents the analysis made to assess the language skills of subjects. A huge majority of subjects are French speakers (mean score : 4.78/5 or 96%) with good English skills (mean score : 4.17/5 or 83%). Even if we ask our questions in both languages, we ask more tasks on websites in English (88%) than on websites in French (12%). Tasks on websites in French are mostly solved (93%) but subjects succeeded well on websites in English too (76%).

Table 4.1: Phase 1: Subjects Language Level analysis.

	in French	in English
Subj. Lang. Level (mean score /5)	4.78	4.17
Subj. Lang. Level (median score /5)	5	4
	Websites in French	Websites in English
Number of Tasks (%)	0.12	0.88
Succeeded Tasks (%)	0.93	0.76

We check if the percentage of succeeded tasks on English websites is the same for subjects with different English skills (cf. Table 4.2). The answer is "yes", since all subjects with English score ≥ 4 or < 4 solved almost the same percentage of tasks (77% and 75%). Consequently, we decide not to differentiate subjects regarding the language skills.

Website Navigability Knowledge

We suppose that these subjects have similar caracteristics as far as the use of the Web is concerned. Since they are all students in Computer Sciences, we assume

Table 4.2: Phase 1: Subjects English Level analysis.

Subjects according to English level	Percentage of Tasks on English Sites	Percentage of Succeeded Tasks (on Eng. Sites)
Engl. Subj. Level < 4	0.16	0.750
Engl. Subj. Level ≥ 4	0.84	0.767

that they have a good experience in website navigability.

4.1.9 Objects

The software artifacts used in the experiment are here 30 real websites. Appendix B presents the list of the 30 websites (section B.1).
These websites have been selected randomly by the mean of a "Random WebSite Machine" (use of `http://www.randomwebsite.net/` and `http://www.randomwebsite.com/`). "The Random Website Machine is very simple. It lets the user surf the web, by sending him a completely random website, selected in a database that contains 4,347,212 websites[4].

Figure 4.1.9 presents the application fields of the 30 selected websites. A more precise description of each website is given in Appendix B (section B.2).

Figure 4.1: Application Fields of the selected websites.

Only websites in English or in French have been retained. We also exclude websites that our program cannot assess (cf. *"Chapter 3; Model Implementation; Limits" section*).
Table 4.3 shows the availability and the distribution of site-level navigation elements among the selected websites.

[4] from `http://www.whatsmyip.org/random_websites/`, 2011-05-02.

Table 4.3: Summary of the 30 selected websites: Availability and distribution of site-level navigation elements.

Availability of	Percentage of websites (%)
Navig. Menu	0.80
Search Engine	0.50
Site Map	0.53
Distribution	Percentage of websites (%)
Menu & Search & Map	0.33
Menu & Search (no map)	0.07
Menu & Map (no search)	0.17
Search & Map (no menu)	0.03
Menu (no search; no map)	0.23
Search (no menu; no map)	0.07
Map (no menu; no search)	0.00
neither of the three	0.10

Finally, we make sure that the selected websites have the same depiction on the main different web browsers (Internet Explorer, Mozilla Firefox, Google Chrome). Figure 4.1.9 gives an example of a website (`http://www.uciprotour.com/`) tested with different web browsers.

4.2 Planning

In this *Planning* section, we define how we conduct the experiment. We precise our objectives and research questions. For each research question, we formulate the null and the alternative hypotheses. The null hypothesis (H_0) is the hypothesis that we want "to reject with as high significance as possible", while the alternative one (H_1) is the hypothesis in favor of which the null hypothesis is rejected.

Then, we select the experiment variables. We want to check a certain number of independent variables. The dependent variable is the studied one (in our case, we study the *navigability* of websites). We want to see what happens when the independent variables on the navigability change from one site to another.

Moreover, we explain how the subjects were selected. We end this section by defining the experiment design and instrumentation.

Figure 4.2: Example of a website tested with 3 different web browsers: Mozilla Firefox 3.6, Internet Explorer 8.0 and Google Chrome.

4.2.1 Hypothesis formulation; Objectives and Research Questions

The experiment first goal is to assess the capacity of the Multi-level Model (distinguishing the page- from the site-level) to predict website navigability. We define two objectives in order to reach our goal. Our two objectives are each divided in two research questions (RQ).

For each research question, we formulate null and alternative hypotheses which are at the basis of the further statistical analysis. We use the collected data from the experiment in order to reject (if possible) the null hypothesis.

The null hypothesis (H_0) is the one that we want to reject. This hypothesis asserts that the only reasons to explain our observations are coincidences.

We define an alternative hypothesis (H_1) with the hope of rejecting H_0. At the end of the experiment, we are able to decide if it is possible to assert H_1 instead of the null hypothesis.

Objective 1

Evaluate whether or not a **page-level assessment** (using the page-level model and the composition model) is able to accurately simulate human judgments.

RQ 1: **Does the choice of a sample of pages impact the result of a page-level model?**

Given selected websites, this question aims to assess if all the webpages of a website statistically have an equivalent navigability level.

RQ1: Null hypothesis statement: H_0: All the webpages of a website statistically have a similar navigability score. Consequently, the choice of a page does not impact the result of a page-level model.

RQ1: Alternative hypothesis: H_1: All the webpages of a website statistically do not have a similar navigability score. Consequently the choice of a page is important for the result of a page-level model.

RQ1: Process: We compute the navigability score of each webpage according to our page-level model. Then, we can see if page-level scores are all similar for a website or not. Thus, assuming that our model is correct (i.e. the page-level model correctly estimates the navigability), we can prove if the *choice of a pages sample* is important or not.

RQ 2: **Can the aggregation of individual page results correspond to human judgments?**

We want to assess a website from the point of view of his webpages navigability level only. Is it possible to find distinct websites categories with the composition model, as assessed by users?

RQ2: Null hypothesis statement: H_0: The "Composition Model" based on the aggregation of individual page results is not able to simulate human judgments about websites navigability assessment.

RQ2: Alternative hypothesis: H_1: The "Composition Model" based on the aggregation of individual page results is able to simulate human judgments about websites navigability assessment.

RQ2: Process: We compute the correlation between navigability scores provided by the "Composition Model" and the subjects estimations of websites navigability. We also analyse if this model is able to distinguish good websites from bad ones, according to human judgments.

Objective 2

Evaluate whether or not the **multi-level model** can produce better estimates.

RQ 3: **Can the multi-level model predict human judgments?**

We want to assess a website thanks to our multi-level model. Is it possible to find distinct websites categories with the multi-level model, as assessed by users?

RQ3: Null hypothesis statement: H_0: The multi-level model is not able to simulate human judgments.

RQ3: Alternative hypothesis: H_1: The multi-level model is able to simulate human judgments.

RQ3: Process: We compute website navigability scores thanks to our multi-level model. We measure the correlation between these navigability scores and human estimates. We also analyse if this model is able to distinguish good websites from bad ones, according to human judgments.

RQ 4: **Is there an importance function (weighting algorithm) that outperforms the others?**

We want to assess the impact of the selected weighting strategy (used in the composition model) on the website navigability.

RQ4: Null hypothesis statement: H_0: There is statistically no difference between websites navigability scores according to the selected weighting strategy.

RQ4: Alternative hypothesis: H_1: Websites navigability scores computed according to different weighting strategies are statistically different. We can find a weighting strategy that outperforms the others.

RQ4: Process: We test several algorithms that assign weight to the different pages of a website. These weights are used to define the importance of a webpage when we compute the global website navigability score. We thus want to know if a given algorithm is better than the others.

4.2.2 Variables selection

This section presents the experiment variables. “Independent variables (or state variables) are the variables that we can control and change in the experiment. The independent variables should have an effect on the dependent variable and must be controllable” [WRH+00].
We define some explanatory variables which are navigability factors not taken into account by the model. However, we ask our subjects to assess them on the experiment questionnaire. They could be viewed as independent variables in future experiments.
There is often only one dependent variable that measures the effect of treatments. We have to measure the dependent variable indirectly. The dependent variable value represents the study results. Our dependent variable is the website’s *navigability*. We measure the quality of navigability at page- and at site-level, according to the quality model we use.
Other factors that may affect the results are discussed in the validity threats section.

Independent variables

Independent variables are supposed to affect the studied quality criterion (*navigability*). According to our model, the input nodes of our BBN affect the navigability node. In the page-level model, input nodes are the assessment scores of the page-level navigation elements. The site-level model uses the page-level results and other input nodes (assessment of site-level navigation elements). All these input sub-criteria (assessment of page- and site-level navigation elements) are our first independent variables.
In the experiment questionnaire, we ask our subjects to assess the site-level navigation elements and some page-level criteria. Our model and subjects assess the independent variables presented in Table 4.4.

Explanatory variables

Here are some other independent variables that the present model does not take into account. However, on the experiment questionnaire, we ask our subjects to

Table 4.4: Independent variables: Input sub-criteria of the Navigability Models.

Page-level Assessment	**Questionnaire** (subject assessment; on a 1 to 4 scale; or ”not tested”)	**Page-level Model**
DisplayTime: Time to display a page	The subject has to estimate the effeciency of downloading a page.	It tallies with the “Down.Speed” node.
BackHome: Coming back to the home page	The subject has to say if he can easily come back to the home page with a proposed link that has to be well visible and clear.	It tallies with the “LinkToHome” node.
VisitedLinks: Recognizing the visited pages	The subject has to estimate if he can easily make the difference between links to the pages that he has already visited and all other links.	The input node “VisitedColour” is used to answer this question.
WhereGo: Knowing where to go after (link quality)	The subject has to estimate how easy it is to know (or to deduce) to which pages or type of pages the links lead him.	Assessing link quality for a subject corresponds to the “TextFeedBack” node (with 3 input sub-nodes: “LinkText”; “MeaningfulURL”; “LinkTitle”).
Locate: Knowing where you are on the website	The subject has to estimate how easy it is to know where he is on the website. Can he see at which level he is in the website (hierarchical indication, current menu button conspicuously presented, etc.).	The input nodes “CurrentPosLabel”, “PathMechanism” (both are sub-nodes of “VisualFeedBack”) and “MeaningfulURL” are used to answer this question.
Site-level Assessment	**Questionnaire** (subject assessment; on a 1 to 4 scale; or ”not tested”)	**Site-level Model**
Navigation Menu; Search Function; Site Map	The subject has to estimate the usability of each site-level navigation element after his session on the website.	Each of these site-level navigation elements is taken into account as an entry node of the site-level Model.

assess these elements about navigability. We want to see how important these factors are with regard to navigability assessment. Subjects have to estimate these elements on a 1 to 4 scale (or “not tested”).

- *Organisation of the navigation elements*: The subject has to judge the organisation quality of the navigation elements. Are these elements well organised, well situated in order to make the search easy and clear (good visibility)?

- *Similarity between the pages*: The subject has to estimate if, moving from one page to another, he can find the navigation elements at the same places and presented similarly.
- *Similarity with other websites*: The subject has to say if the presentation and organisation of the navigation elements are similar to what he usually sees on the web.

On the experiment questionnaire, we also ask our subjects to assess each navigation session as a whole.

- *Estimation of time* : On a scale ”Instantanneous ; relatively short ; relatively long”, the subject has to estimate the time he needed to solve the task (solved or not).
- *Finding easily the target information*: The subject has to say if it is usually easy to solve the task, taking into account the simplicity level of the task and the helpfulness of the available navigation elements to perform the task (on a 1 to 4 scale; or ”not tested”).
- *Satisfaction*: The subject has to say if he is personally satisfied after his navigation session on the website (on a 1 to 4 scale; or ”not tested”).

Finally, we discuss other factors (such as subjects, websites and tasks selection, other experiment controls, etc.) in the “Validity Assessment” section.

Dependent variable

Dependent variables are influenced by the independent variables. Dependent variables are not directly measurables. They must be derived from the hypothesis. Our dependent variable is the site’s **navigability**. This variable is estimated through our model. It modelizes navigability as a probability, thanks to its input metrics (the first independent variables above-mentioned). First, we compute website navigability thanks to a page-level assessment (using the page-level model and the composition model). Second, we compute the dependent variable through our entire multi-level model (including the site-level model).
On the experiment questionnaire, we ask our subjects to give a *global estimation*

of navigability: a global score for the website navigability (on a 1 to 10 scale) when the subject finishes his task. We want to analyse if subjects estimates correlate with navigability scores computed by our composition model on the one hand (page-level assessment), and with site-level navigability scores on the other hand (cf. Research Questions 2 and 3).

4.2.3 Experiment design

Random distribution

The websites are randomly selected in order to represent the web and its huge variety as best as possible.
For each website, we write 3 different tasks. We have 30 different websites and therefore 90 different tasks. Each subject receives 5 randomly chosen tasks. The only constraint is that a subject does not assess one website more than one time. We need at least 18 subjects (18*5=90) to perform the experiment. During the first phase, 22 subjects (18 + 4 extra) take part in the experiment, that is why 20 tasks (4*5) are completed twice. Tables 4.5 and 4.6 show the tasks distribution.

Blocking technique

Before the operation phase of the experiment, we define two potential ways of creating "blocks" among the subjects.
Language skills: We check the language skills of the subjects in English and French. We have already explained why we decided not to differentiate subjects regarding language skills (section 4.1.8). *Web use level*: In the ideal case of a general approach with subjects from everywhere, we have to group the subjects into different "blocks":

- Expert: subjects that have a current and everyday use of the Web.
- Intermediate: subjects that use the Web more occasionnaly.
- Novice: subjects that discover the Web and how to use it.

In our experiment, we assume that all our subjects belong to the first "Experts" block . De facto, the subjects are all students or researchers in Computer Science and are accustomed to using websites.

Table 4.5: Phase 1: Tasks distribution (according to subjects).

subj01	siteID08 id23	siteID01 id02	siteID12 id34	siteID11 id30	siteID03 id06
subj02	siteID10 id28	siteID15 id42	siteID16 id46	siteID13 id36	siteID19 id55
subj03	siteID11 id31	siteID05 id13	siteID20 id57	siteID06 id15	siteID14 id41
subj04	siteID05 id14	siteID04 id10	siteID18 id52	siteID17 id48	siteID10 id27
subj05	siteID15 id43	siteID17 id49	siteID11 id32	siteID20 id58	siteID13 id38
subj06	siteID02 id05	siteID20 id59	siteID09 id26	siteID10 id29	siteID14 id39
subj07	siteID01 id00	siteID17 id50	siteID16 id47	siteID07 id18	siteID03 id08
subj08	siteID08 id22	siteID04 id11	siteID12 id35	siteID06 id16	siteID18 id51
subj09 + subj10	siteID01 id01	siteID14 id40	siteID13 id37	siteID09 id25	siteID19 id54
subj11	siteID15 id44	siteID03 id07	siteID05 id12	siteID07 id19	siteID09 id24
subj12	siteID21 id60	siteID30 id89	siteID23 id66	siteID26 id76	siteID16 id45
subj13 + subj14	siteID24 id70	siteID27 id78	siteID25 id73	siteID30 id88	siteID04 id09
subj15 + subj16	siteID23 id68	siteID29 id84	siteID26 id77	siteID24 id71	siteID08 id21
subj17	siteID19 id56	siteID23 id67	siteID21 id62	siteID25 id74	siteID18 id53
subj18	siteID25 id72	siteID24 id69	siteID22 id65	siteID29 id85	siteID02 id03
subj19 + subj20	siteID22 id63	siteID21 id61	siteID30 id87	siteID28 id83	siteID07 id20
subj21	siteID26 id75	siteID28 id81	siteID29 id86	siteID27 id80	siteID06 id17
subj22	siteID28 id82	siteID22 id64	siteID27 id79	siteID02 id04	siteID12 id33

Balancing technique

During the experimment each subject has to do 5 tasks. Each site is at least evaluated by 3 different subjects who all perform a different task on the website. We explained above (cf. Random distribution) that some websites were assessed

Table 4.6: Phase 1: Tasks distribution (according to websites).

site	nbSubj	task1	task2	task3
site01	4	subj07	subj09 + subj10	subj01
site02	3	subj18	subj22	subj06
site03	3	subj01	subj11	subj07
site04	4	subj13 + subj14	subj04	subj08
site05	3	subj11	subj03	subj04
site06	3	subj03	subj08	subj21
site07	4	subj07	subj11	subj19 + subj20
site08	4	subj15 + subj16	subj08	subj01
site09	4	subj11	subj09 + subj10	subj06
site10	3	subj04	subj02	subj06
site11	3	subj01	subj03	subj05
site12	3	subj22	subj01	subj08
site13	4	subj02	subj09 + subj10	subj05
site14	4	subj06	subj09 + subj10	subj03
site15	3	subj02	subj05	subj11
site16	3	subj12	subj02	subj07
site17	3	subj04	subj05	subj07
site18	3	subj08	subj04	subj17
site19	4	subj09 + subj10	subj02	subj17
site20	3	subj03	subj05	subj06
site21	4	subj12	subj19 + subj20	subj17
site22	4	subj19 + subj20	subj22	subj18
site23	4	subj12	subj17	subj15 + subj16
site24	5	subj18	subj13 + subj14	subj15 + subj16
site25	4	subj18	subj13 + subj14	subj17
site26	4	subj21	subj12	subj15 + subj16
site27	4	subj13 + subj14	subj22	subj21
site28	4	subj21	subj22	subj19 + subj20
site29	4	subj15 + subj16	subj18	subj21
site30	5	subj19 + subj20	subj13 + subj14	subj12

twice in the first experiment phase.

Our replication phase uses a perfect balanced design: each subject performs 5 tasks and each website is assessed by 3 subjects (We work with 40 websites: the same 30 websites of the first phase and 10 additional sites which we have not finally taken into account in the analysis).

4.2.4 Instrumentation

We present three different types of instruments for the experiment. The objects, the guidelines, and the measure instruments.

Objects

The objects of the experiment are the 30 websites. Appendix B presents the list of the 30 websites (section B.1). These websites are analysed by the subjects and also by our Java program, according to the multi-level model.

Guidelines

Each subject who takes part in the experiment receives an explanation document with the goal of the experiment, the rules, and the definitions used in the online questionnaire. Appendix B contains a copy of this document (section B.3).

Measure instruments

The experiment has two faces. The first one is the assessment of websites by users. For each subject, this step consists in subscribing to the experiment, reading the explanation about the experiment, and finally completing the online questionnaire. Appendix B contains a copy of the online questionnaire (section B.3.2). Then, we gather the results on separate spreadsheets for each task.
The other face concerns the measurement executed by our related tool. Each time, the measurement takes place when the subject is completing the task on a website (cf. *"Validity Assessment; Conclusion validity threats; Reliability of measures"*).

4.3 Validity Assessment of the Experiment

We have to face different threats to validity that occur in every experiment process. Here are the main validity threats we notice and how (and to what extent) we manage them.

4.3.1 Conclusion validity threats

"Threats to conclusion validity are concerned with issues that affect the ability to draw the correct conclusion about relations between the treatment and the outcome of an experiment. These issues include, for example, choice of statistical tests, choice of sample sizes, care taken in the implementation and measurement of an experiment" [WRH+00].

Reliability of measures: The navigability scores produced by our model are computed thanks to our related tool. We previously presented some limits of our Java program (Chapter 3; section 3.4). "Reliability of measures" threat concerns the process used to download and assess a site. Our tool is able to compute navigability scores for all the selected websites, though our crawler might miss pages. However, for huge websites, we compute navigability scores for a portion of the site (one hundred pages) and for the complete site (hundreds, sometimes one thousand of pages). In this case, we do not see any significant difference in navigability scores. Therefore, we choose to compute the navigability score of a website working on its first 200 pages. Moreover, we systematically check that our program accurately detects the site-level navigation elements.

Regarding our experiment questionnaire, results are directly gathered on a spreadsheet and then grouped by websites. We check some elements to be sure that subjects answer seriously and without bias. For each task, we check if the subject knows the website or not. We also check the language skills of subjects (section 4.1.8). After the task, the subject must indicate if he succeeded in solving the asked task or not. Then, he is asked to give the task answer (the URL of the targeted page or, for example, the price of the searched item or a precise piece of information, etc.). If the subject has not succeeded in performing the task, he has to explain why (for example, the task was too difficult to understand, or a navigation element was missing on the site or has poor quality, etc.).

Random irrelevancies in experimental setting: Some elements outside the experimental setting may disturb the results (e.g., noise, sudden interrupt in the experiment). We assert that such elements do not take place during supervised sessions. Regarding the remote-controlled sessions (subjects that peform the experiment at home), we cannot assert that no elements disturbed the session. However, we check the successive timestamps between the different tasks and we can assert that each experiment is "all in one piece".

Random heterogeneity of subjects: We previously explained the subjects selection (section 4.1.8). Our samples are composed of both students and researchers in Computer Science. We choose an homogeneous group of subjects to avoid that navigability judgments depend on the web experience and knowledge. We do not have enough subjects to study navigability assessment with different categories of subjects (for example: novice, intermediate, and expert according to web use level).

Data manipulation: We pay attention to the way we manipulate data in the study. We calculate the mean of subjects estimates about the site's navigability and about the navigation elements. "The use of the mean is possible, because in both 1-to-10 and 1-to-5 scale cases, we consider that the measured variables have an interval scale. This is possible because we assigned a meaning to the limit values (1 and respectively 5 or 10), but not to the intermediate values. This prevents from considering the variables as having a likert (ordinal) scale" [MSBB10]. Most tests and correlations (Pearson) are parametric, requiring a normal distribution. We use a Kolmogorov-Smirnov test, which indicates that the data produced by our models and subject judgments is normally distributed. We use correlation tests "to estimate the strength of the relation between the variables" [MSBB10]. We present the data manipulation in the next chapter.

4.3.2 Internal validity threats

"Threats to internal validity concern issues that may indicate a causal relationship, although there is none. Factors that impact on the internal validity are how the subjects are selected and divided into different classes, how the subjects are treated and compensated during the experiment, if special events occur during the experiment etc. All these factors can make the experiment show a behavior that is not due to the treatment but to the disturbing factor" [WRH+00].

Bias Avoidance: We try to avoid exerting any influence on subjects when they first assess the website. Subjects are free to explore the site, with no intervention on our part. We ask them to assess the website navigability before describing the different mechanisms used. Thus, the subjects do not consider the presence of these navigation mechanisms before assigning a score to the website.

Maturation: To avoid a maturation threat (i.e. learning and fatigue effects), the list of websites to assess is randomly distributed among the subjects. Websites are thus assessed at random and in different orders (cf. section 4.2.3).

History: In the experiment, the same websites are assessed by our model and by different subjects. Automated measurement and subject session take place at the same moment (at least on the same day). We check that our model computes identical navigability scores when it works several times, sometimes with a few days gap. If navigability scores from our model are statistically different for the same website, we keep the mean score.

Instrumentation: As presented in section 4.2.4, subjects receive an experiment explanation with their questionnaire. This explanation allows each subject to have the same definitions of our navigability concepts. Even if it is difficult to be completely sure that everyone has well read and understood the experiment rules and definitions, some elements reassure us. The sended mail to a subject clearly mentions the linked explanation document, the most important remarks are summed up on the questionnaire. Moreover, about half of the subjects has perform the experiment in the laboratory where we work. Consequently, we were sure that the subject read the document. If necessary, we can answer his questions or clarify some points. Regarding the model assessment, we have already explained what we check (cf. "History" just above and "Reliability of measures" in the previous "Conclusion validity threats" section).

Websites Selection: Section 4.1.9 presents how the websites are randomly selected.

Tasks Selection: We ask a variety of tasks that could be accomplished by means of different mechanisms. Three tasks are defined for each selected website. These tasks consist in searching a precise piece of information on the website, like the price of a particular item on a commercial website, a phone number, or a target webpage, etc. Finding tasks with similar difficulty level is not an obvious thing (different websites, small- or large-sized, in different fields). We tried to choose tasks related to the core business of the website (for example, to find an item price on a retail website).

4.3.3 Construct validity threats

"Threats to construct validity refer to the extent to which the experiment setting actually reflects the construct under study" [WRH+00].

Mono-operation Bias: We must avoid to use "a single version of a program in a single place at a single point in time"[5]. We have already explained that we use 30 different websites (randomly selected) and we assess navigability at multiple levels. Moreover, our experiment comprises two phases (first phase at the University of Montréal and replication at the University of Namur). Consequently, navigability automated assessment takes place simultaneously for each experiment phase.

Mono-method Bias: We have to measure navigability and provide different navigability measures. Our model assesses navigability at page- and site-level. We also test several weighting strategies through our composition model.
In the experiment questionnaire, we ask our subjects to assess navigability. They also have to assess page- and site-level navigation elements. Finally, we ask them to assess the task as a whole (ease of finding the targeted piece of information, satisfaction). We can check that the estimates of each subject are consistent.

Evaluation Apprehension: We reassure our subjects that they are not marked and that their anonymity will be preserved. We also explain our subjects that the time they take to perform a task is not recorded.

4.3.4 External validity threats

"Threats to external validity concern the ability to generalize experiment results outside the experiment setting. External validity is affected by the experiment design chosen, but also by the objects in the experiment and the subjects chosen. There are three main risks: having wrong participants as subjects, conducting the experiment in the wrong environment and performing it with a timing that

[5]From `http://www.socialresearchmethods.net/kb/consthre.php`; (Date of access 2011-04-18).

affects the results" [WRH+00].

Web Representativeness: We try to represent the Web as best as possible and select random websites. However, we realize that it is impossible "to epitomize the Web" with only 30 websites. Possible future work consists in testing other types of websites.

Web Users Representativeness: Concerning the Web use level of subjects, only one category is taken into account for the experiment: expert. We would like to investigate how different subjects (e.g. who are not studying Computer Science) might respond differently.

4.4 Operation

We have already explained how the experiment concretely works (cf. How does it work? section). We just remind the three main operation steps of our experiment.

4.4.1 Preparation

First, we finish our model implementation (cf. Related Tool) that we are going to use to assess websites navigability.
Second, we prepare our experiment: we randomly select 30 websites, we find the navigation tasks, we write the experiment explanation and questionnaire (French and English versions) and we put it on line (cf. Experiment Questionnaire). We choose the experiment design too (cf. Experiment design section).
Third, students and researchers in Computer Science (from the University of Montréal for the first experiment phase, from the University of Namur for the replication) are invited by email to take part in the experiment. This invitation briefly explains the experiment goal and the way it works. Then, all those interested receive the complete explanation and the link to the experiment questionnaire. These persons are the subjects.

4.4.2 Execution

As soon as a subject receives the link to his online questionnaire, he is free to perform the experiment. For each phase, we also organize a lab session where several subjects perform the experiment at the same time. It allows us to be sure

subjects understand well the experiment goal and terms. We gather questionnaire results on an individual spreadsheet for each task.
At the same time, we measure websites navigability thanks to our Java program (cf. Chapter 3; Related Tool).

4.4.3 Data validation

Before analysing the results, we check some questionnaire answers: language skills (cf. Subjects section), prior knowledge of the website and tasks answers. We explain data manipulation and results in the next chapter.

Chapter 5

Experiment Results and Discussion

In the third chapter, we presented a quality model that aims to assess website navigability. Like other quality models, our Multi-level Model intends to reproduce human judgment. In order to achieve it, we set up an experiment (cf. *Chapter 4*) that aims to validate our Multi-level model. In this experiment, we compare human assessments about websites navigability with the navigability scores computed by our program for the same websites.
Our main goal in this chapter is to see if the Multi-level Model is able to accurately simulate human judgment and if it is better than a single page-level model. First, we try to answer the research questions introduced in Chapter 4. Then, we present some other results of our study. We analyse the questionnaire answers about page- and site-level navigation elements.
We finally discuss how our model can be used to improve website navigability thanks to an important property of Bayesian Belief Networks (BBNs).

5.1 Research Questions Results

In this section, we present the main results of the experiment. We tend to answer the research questions introduced in Chapter 4 (cf. section 4.2).

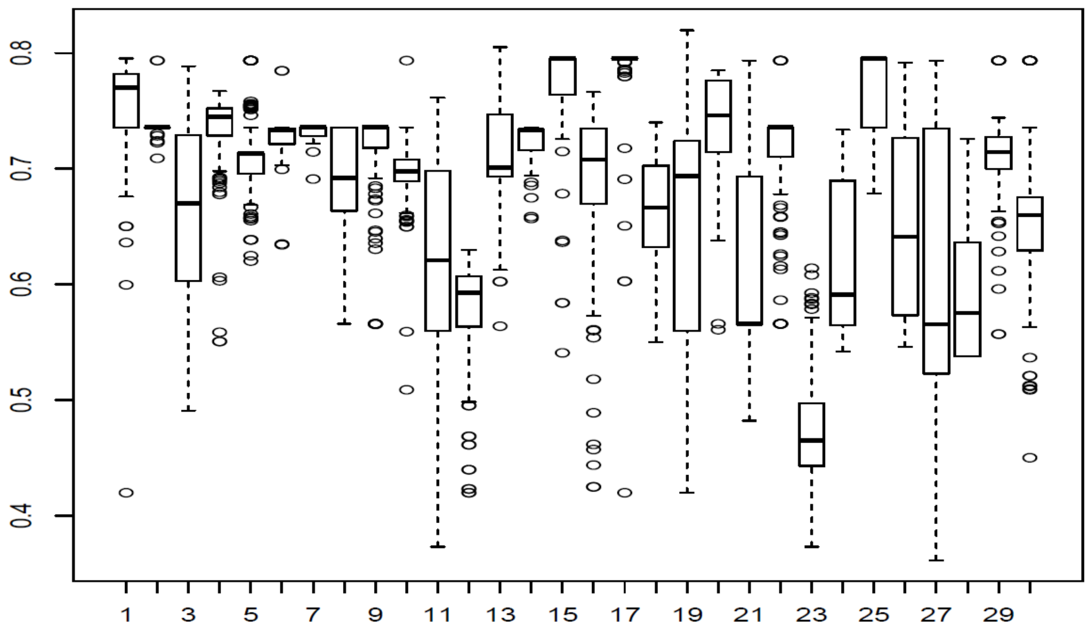
5.1.1 RQ1 Results - Impact of the choice of a page

***RQ1:* Does the choice of a page impact the result of a page-level model?**

When we assess the selected websites, we automatically assess the webpages of these sites. We concretely apply the page-level model (cf. Chapter 3; section 3.3.1) to the different pages downloaded from the sites studied.

Figure 5.1 shows the variability of the webpages assessments as boxplots[1]; each of the 30 websites is presented separately on the x-axis. The y-axis gives the navigability scores (probability of the webpage navigability to be "good"). Regarding the variability of the webpages scores computed by our model, we can answer positively to RQ1. We observe that few websites have pages with stable

Figure 5.1: Boxplots of page-level navigability scores for each website.



navigability scores. Many sites have a large number of outliers (identified as circles outside the whiskers). These outliers mean that, if the model is accurate, any user travelling through one of these pages would probably have an unrepresentative impression of the "global" quality. Therefore, we believe that webpage

[1]A boxplot or *Box and Whisker Plot* is "a graphical representation of dispersion of the data. The graphic represents the lower quartile (Q1) and upper quartile (Q3) along with the median. The median is the 50th percentile of the data. A lower quartile is the 25th percentile, and the upper quartile is the 75th percentile. The upper and lower fences usually are set a fixed distance from the interquartile range (Q3 − Q1). Any observation outside these fences is considered a potential outlier. Even when data are not normally distributed, a box plot can be used because it depends on the median and not the mean of the data" [Wal06].

choice is important.

5.1.2 RQ2 Results - Page-level Analysis

***RQ 2:* Can the aggregation of individual page results (through our composition model) correspond to human judgments?**

We tested how the model could be used to assess the quality of the sites by averaging the scores of individual pages. First, we check the normality assumption. Then, we use several tests to answer RQ2.

Normality assumption

Most tests and correlations (Pearson) are parametric, requiring a normal distribution. We thus check if the data is normally distributed. We use a Kolmogorov-Smirnov (K-S) test to evaluate the normality assumption. We briefly remind[2] how we can explain the results of this test. We use it to compare a sample with a reference probability distribution (one-sample K–S test) which is the normal distribution in our case. Our data contains the scores produced by our models and subject judgments. K-S test results are presented in Figure 5.2.

Figure 5.2: Phase 1: Page-level analysis: Normal distribution check: Kolmogorov-Smirnov test results.

One-Sample Kolmogorov-Smirnov Test

		scorePages SM	scorePages VP	MeanScore Subj	MedianScore Subj
N		30	30	30	30
Normal Parameters[a,,b]	Mean	,682695094	,680472027	5,687	5,78
	Std. Deviation	,0671491431	,0754665896	2,1064	2,690
Most Extreme Differences	Absolute	,164	,217	,171	,195
	Positive	,095	,148	,105	,115
	Negative	-,164	-,217	-,171	-,195
Kolmogorov-Smirnov Z		,899	1,188	,936	1,068
Asymp. Sig. (2-tailed)		,394	,119	,345	,204

a. Test distribution is Normal.

b. Calculated from data.

The *Asymp. Sig. (2-tailed)* value is also known as the p-value. The p-value tells us the probability of getting the results we got if the null were actually true

[2]According to J. Mirabella, *Hypothesis testing with SPPS*, 2006 (`http://www.drjimmirabella.com/ebook/excerpt%20from%20Hypothesis%20Testing%20with%20SPSS%20ebook%20%28Jim%20Mirabella%29.pdf`); Date of access 2011-07-14.

(i.e., it is the probability we would be wrong if we rejected the null hypothesis). The hypotheses for this test of normality are:
H_0: The distribution of the sample is normal.
H_a: The distribution of the sample is not normal.
If the p-value is less than .05, we reject the normality assumption. If the p-value is higher than .05, there is insufficient evidence to suggest the distribution is not normal (meaning that we can proceed with the assumption of normality).
Since each p-value is higher than .05, there is no reason to doubt the distribution is normal, so we can proceed with the correlations and the t-test.
Appendix C contains the analysis of the replication phase. Section C.3.1 presents the results of the page-level analysis. Figure C.1 gives the K-S test results for the replication phase.

Correlations

Figure 5.3 presents the correlation between navigability scores computed for webpages and human judgments of navigability. Each webpage score is computed by the page-level model. The pages scores are then aggregated using Visit Probability (VP) and Simple Mean (SM) strategies. These scores are compared with human judgments about navigability (mean and median navigability score, as estimated by subject for each website).

We can answer "no" given the obtained results. Our correlation analysis indi-

Figure 5.3: Page-level analysis: correlations between navigability scores (computed thanks to a page-level model) and human judgments.

		scorePages VP	scorePages SM	MeanScore Subj	MedianScore Subj
scorePagesVP	Pearson Correlation	1	,944**	,372*	,381*
	Sig. (2-tailed)		,000	,043	,038
	N	30	30	30	30
scorePagesSM	Pearson Correlation	,944**	1	,388*	,390*
	Sig. (2-tailed)	,000		,034	,033
	N	30	30	30	30
MeanScoreSubj	Pearson Correlation	,372*	,388*	1	,956**
	Sig. (2-tailed)	,043	,034		,000
	N	30	30	30	30
MedianScoreSubj	Pearson Correlation	,381*	,390*	,956**	1
	Sig. (2-tailed)	,038	,033	,000	
	N	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

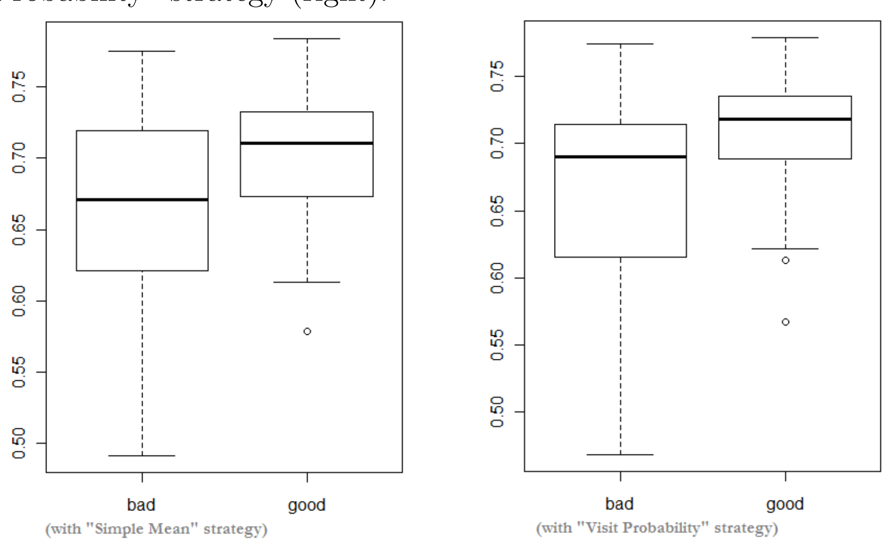
*. Correlation is significant at the 0.05 level (2-tailed).

cates that the navigability scores are significantly correlated to the judgments of our subjects but the correlation is not very strong (0.39). Figure C.2 gives the same correlation for the replication phase. This correlation is even lower (0.21) and not significant.

Compared distribution

Furhermore, the composition model cannot clearly distinguish a good site from a bad site. We assert it in view of the distribution of predicted values for the good and bad sites. We consider that “bad websites” are the ones poorly estimated by subjects (subjects mean score ≤ 5). “Good websites” are the ones getting an average score of five and more (subjects mean score > 5). We choose the value 5 because it corresponds not only to the halfway point in our scale, but also because it splits the corpus in two almost equal sized parts.
We draw two boxplots with the following parameters: Figure 5.4 represents the distribution of bad and good websites. Here, the scores are computed as follows: each webpage score is computed according to our page-level model. Then, they are aggregated according to the Simple Mean strategy (the global score for webpages is the mean of each webpage navigability score) or the Visit Probability strategy. Figure C.3 shows the same with the results of the replication phase.

Figure 5.4: Compared distribution of scores for good and bad sites by means of an aggregated page-level model; with “Simple Mean” strategy (left) and with “Visit Probability” strategy (right).



A population test (using a t-test since the data is normally distributed) indicates that we cannot assert that the difference in means is significant. The t-test assesses whether the means of two groups are statistically different from each other. We just remind how the t-test works[3] and how we do it here:

1. The t-test assumes that the distribution is normal (parametric data) and that underlying variances are equal. In our case, the data is normally distributed and we use the Welch's t test, that is an adaptation of Student's t-test, intended for two samples having possibly unequal variances.

2. The basic principle is to test the null hypothesis, which states that the means of the two groups are equal. Here, we have two groups of websites: "good" and "bad".

    - The "good websites" are those that subjects estimate navigability above 5/10 (mean and median score). We take the navigability scores computed by our composition model for these websites, according to Simple Mean (SM) and Visit Probabibility (VP) strategies.

    - It is the same for "bad websites" which subjects estimate badly (below 5/10). Once again, we keep the page-level navigability scores (computed by our composition model) of these websites.

3. The null hypothesis: $H_0 : \mu_{good} = \mu_{bad}$
   The alternative hypothesis: $H_1 : \mu_{good} \neq \mu_{bad}$
   Where μ is the mean score of each group.

4. To test the significance, we set a risk level (called the alpha level). We specify the α level: $\alpha = .05$

5. The test is done. Different values are computed:

    - **t-value** is a ratio: "difference between group means" / "variability of groups (or standard error of difference between group means)". The t-value will be positive if the first mean is larger than the second and

[3]From http://changingminds.org/explanations/research/analysis/t-test.htm and http://www.socialresearchmethods.net/kb/stat_t.php (Date of access 2011-07-09).

negative if it is smaller.

- **df** are the degrees of freedom.

- **p-value**: the significance. If this value is lower than or equal to our α level for this test, we can reject the null hypothesis. If it is higher than .05, we cannot reject H_0.

Table 5.1 presents t-tests results. For example, the first t-test gives p-value = 0.132 ; that is not lower than or equal to .05, so we cannot reject the null hypothesis. In each case, we cannot assert that the difference in means is significant, because we always fail to reject the null hypothesis. It is the same for the replication results (cf. Table C.6).

Table 5.1: Page-level analysis : Population test.

t-test param.	t	df	p-value	Mean (good)	Mean (bad)
scorePagesSM by SubjMeanScore > 5	-1.5609	23.435	0.1320	0.701	0.662
scorePagesSM by SubjMedianScore > 5	-2.0324	24.965	0.0529	0.706	0.659
scorePagesVP by SubjMeanScore > 5	-1.4173	20.867	0.1712	0.699	0.659
scorePagesVP by SubjMedianScore > 5	-1.8307	22.638	0.0803	0.705	0.656

Therefore, we believe that we cannot simply aggregate page-level results. We do the same tests using navigability scores computed by our entire Multi-level Model (including the site-level model) to see if our multi-level model is able to successfully ditinguish between good and bad sites (cf. next RQ).

5.1.3 RQ3 Results - Site-level Analysis

RQ3: Can the multi-level model predict human judgments?
We apply our multi-level model to the studied websites. Similarly to RQ2, we first check the normality assumption in order to calculate the correlations and to proceed with the t-test.

Normality assumption

Before the correlations and the population test (using a t-test), we check again if our data is normally distributed. Figure 5.5 gives the normality test (using a K-S test) results. We remind this test principle just above (cf. RQ2). Since each p-value (called *Asymp. Sig. (2-tailed)* in the table) is higher than .05, there is no reason to doubt the distribution is normal, so we can proceed with the t-test. Appendix C presents the site-level analysis for the replication phase (cf. section C.3.2). Figure C.4 gives the K-S test results for the replication phase.

Figure 5.5: Site-level analysis: Normal distribution check: Kolmogorov-Smirnov test results.

One-Sample Kolmogorov-Smirnov Test

		scoreSiteSM	scoreSiteVP	MeanScore Subj	MedianScore Subj
N		30	30	30	30
Normal Parameters[a,b]	Mean	,760942288	,758908910	5,687	5,78
	Std. Deviation	,1405150110	,1475899464	2,1064	2,690
Most Extreme Differences	Absolute	,160	,170	,171	,195
	Positive	,146	,154	,105	,115
	Negative	-,160	-,170	-,171	-,195
Kolmogorov-Smirnov Z		,877	,930	,936	1,068
Asymp. Sig. (2-tailed)		,426	,352	,345	,204

a. Test distribution is Normal.

b. Calculated from data.

Correlations

The correlation between navigabilty scores computed by our model and human judgements (cf. Figure 5.6) is relatively strong (0.72, compared to 0.37 using the page-level model) and significant. Figure C.5 gives the same results for the replication phase: the correlation is as strong (0.74) and significant at the 0.01 level.

Compared distribution

The model is able to successfully ditinguish between good and bad sites (cf. Figure 5.7). It was not possible with the page-level model (cf. previous question). As previously explained (cf. RQ2), we assert it in view of the distribution of predicted values for good and bad sites. We assume that "bad" websites are poorly estimated by subjects (subjects mean score ≤ 5) and "good" ones have scores > 5.

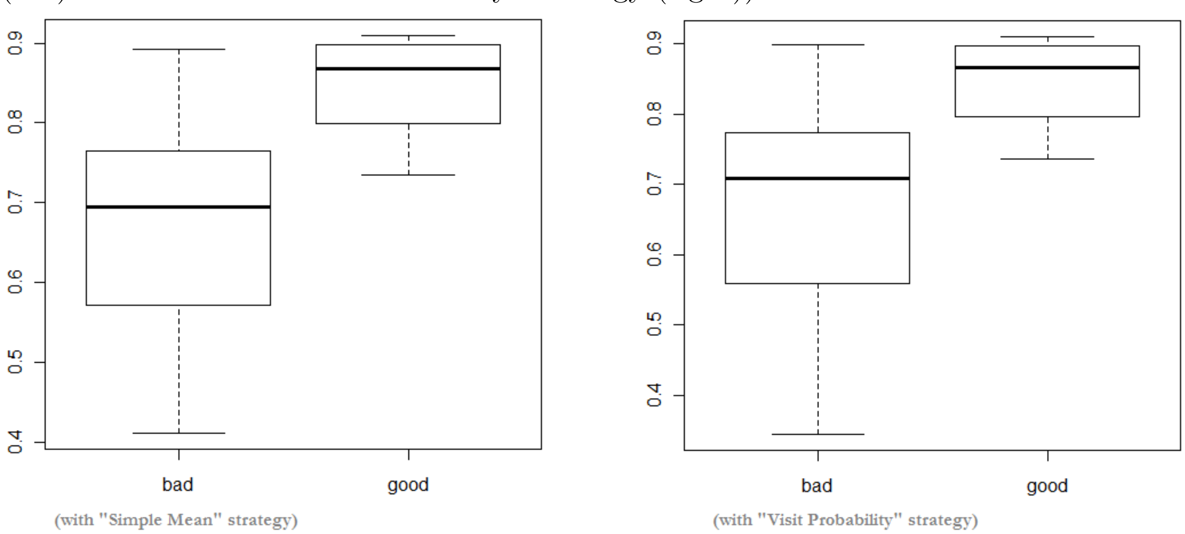
Figure 5.6: Site-level analysis: correlations between websites navigability scores and human judgments.

		scoreSiteVP	scoreSiteSM	MeanScore Subj	MedianScore Subj
scoreSiteVP	Pearson Correlation	1	,996**	,715**	,670**
	Sig. (2-tailed)		,000	,000	,000
	N	30	30	30	30
scoreSiteSM	Pearson Correlation	,996**	1	,737**	,686**
	Sig. (2-tailed)	,000		,000	,000
	N	30	30	30	30
MeanScoreSubj	Pearson Correlation	,715**	,737**	1	,956**
	Sig. (2-tailed)	,000	,000		,000
	N	30	30	30	30
MedianScoreSubj	Pearson Correlation	,670**	,686**	,956**	1
	Sig. (2-tailed)	,000	,000	,000	
	N	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

We draw the two same boxplots but we now use the scores computed by our multi-level model. This navigability model takes into account the site-level navigation elements and the webpages score (computed according to "Simple Mean" or "Visit Probability" strategies). Figure C.6 presents the same distribution with the results of the replication phase.

Figure 5.7: Compared distribution of scores for good and bad sites using the multi-level model (webpages score computed with the "Simple Mean" strategy (left) and with the "Visit Probability" strategy (right)).



We can assert that the difference in means (between the same two groups: "bad sites" and "good sites", using the websites navigability scores computed by our multi-level model) is significant thanks to a population test (we previously explained the t-test working and interpretation, cf. RQ2).
Table 5.2 presents t-tests results. For example, the last t-test of this table gives p-value = 0.001; that is less than .05, so we can reject the null hypothesis. In each case, we can assert that the difference in means is significant, because we can always reject the null hypothesis (each p-value is here $\leq 0.05\_(\alpha)$). It is the same with the t-test results for the replication phase (cf. Table C.7). The multi-level model is thus able to distinguish bad websites from good ones.

Table 5.2: Site-level analysis : Population test.

t-test param.	t	df	p-value	Mean (good)	Mean (bad)
scoreSiteSM by SubjMeanScore > 5	-4.5471	16.978	0.0003	0.847	0.662
scoreSiteSM by SubjMedianScore > 5	-4.1172	18.718	0.0006	0.846	0.676
scoreSiteVP by SubjMeanScore > 5	-4.272	16.482	0.0005	0.847	0.658
scoreSiteVP by SubjMedianScore > 5	-3.9013	18.215	0.001	0.845	0.673

We therefore conclude that the multi-level model can simulate human judgments and outperforms the page-level model when its results are averaged (by means of the composition model).

5.1.4 RQ4 Results - Choice of a weighting strategy

RQ4: Is there an importance function (weighting algorithm) that outperforms the others?

We compute websites navigability scores by different ways, using some weighting algorithms in order to predict the navigability score for the *Pages*. We explained these algorithms in Chapter 3 (*Multi-level Approach* section; *Composition Model*; cf. 3.3.2).
To answer RQ4, we compare the correlation coefficient between human judgments and the navigability score computed according to each importance algorithm. Table 5.3 presents these correlations. We observe that all strategies are significant

Table 5.3: Correlations between scores computed by the Multi-level Model (using different weighting algorithms) and human judgments about navigability (subjects mean score).

Weighting Algo.	correl. coef.	sig.
SimpleMean	0.74	0.000
VisitProbability	0.72	0.000
PageRank	0.74	0.000
Betweeness	0.69	0.000
HITS (hub)	0.74	0.000

and most results are indistinguishable from one another. Therefore, we cannot assert that a particular importance function will produce better results than another. We would recommend using a simple strategy like an equal weight strategy in a composition model.

Fore more details, we present the correlations results for the first experiment phase in Figure 5.8 and for the replication phase in Figure C.9.

Figure 5.8: Experiment first phase: : Correlations between navigability scores for 30 websites (according to different methods to aggregate webpages scores), and with mean and median navigability scores estimated by subjects.

Correlations (Experimentation - 1st phase)

		Visit Probability	Simple Mean Of Pages Scores	Weighted-NI-Paths	Betweenness Centrality	Random-Walk Betweenness	PageRanks	HITS_hub	Markov Centrality	K-Step Markov	MEAN_subjects	MEDIAN_subjects
Visit Probability	Pearson Correlation	1	,996**	,997**	,982**	,999**	,996**	,996**	,996**	,998**	,715**	,670**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Simple Mean Of Pages Scores	Pearson Correlation	,996**	1	1,000**	,981**	,997**	1,000**	1,000**	1,000**	,994**	,737**	,686**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Weighted-NI-Paths	Pearson Correlation	,997**	1,000**	1	,982**	,998**	1,000**	1,000**	1,000**	,996**	,733**	,684**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Betweenness Centrality	Pearson Correlation	,982**	,981**	,982**	1	,984**	,981**	,981**	,981**	,981**	,687**	,631**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Random-Walk Betweenness	Pearson Correlation	,999**	,997**	,998**	,984**	1	,997**	,997**	,997**	,999**	,721**	,678**
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
PageRanks	Pearson Correlation	,996**	1,000**	1,000**	,981**	,997**	1	1,000**	1,000**	,994**	,737**	,686**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
HITS_hub	Pearson Correlation	,996**	1,000**	1,000**	,981**	,997**	1,000**	1	1,000**	,994**	,737**	,686**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Markov Centrality	Pearson Correlation	,996**	1,000**	1,000**	,981**	,997**	1,000**	1,000**	1	,994**	,737**	,686**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
K-Step Markov	Pearson Correlation	,998**	,994**	,996**	,981**	,999**	,994**	,994**	,994**	1	,709**	,666**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
MEAN_subjects	Pearson Correlation	,715**	,737**	,733**	,687**	,721**	,737**	,737**	,737**	,709**	1	,956**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000
	N	30	30	30	30	30	30	30	30	30	30	30
MEDIAN_subjects	Pearson Correlation	,670**	,686**	,684**	,631**	,678**	,686**	,686**	,686**	,666**	,956**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	
	N	30	30	30	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

5.2 Questionnaire results

We only use the navigability scores estimated by users to answer our research questions. We describe here other relevant results of our questionnaire. Our goal is to discover which other factors a quality model should take into account in order to assess navigability. We analyse several descriptive questions included in the experiment questionnaire.

The first part of the questionnaire focuses on site-level navigation mechanisms. Then, we ask to our subjects about the key page-level factors that impact the perceived navigability of the sites.

5.2.1 Questionnaire Analysis: Site-level questions

Among the 30 assessed websites, 24 have a Navigation Menu (80%), 15 have an intern Search Function (50%) and 16 have a Site Map (53%).

We ask our subjects to assess these site-level navigation mechanisms. Figure 5.9 presents the correlations between the average estimates about the site-level navigation elements (when available) and the navigability scores (mean and median) assessed by subjects.

Figure 5.9: Phase 1: Questionnaire analysis - Correlations between the estimates about the site-level navigation elements and the navigability estimates.

Phase 1

Correlations

		menu	search	siteMap
mean	Pearson Correlation	,752**	,557*	,473*
	Sig. (2-tailed)	,000	,025	,035
median	Pearson Correlation	,721**	,592*	,518*
	Sig. (2-tailed)	,000	,016	,019

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

The "menu" estimates are significantly correlated (at 0.01 level) with the navigability assessments. The estimates about the other two navigation elements are also correlated with the navigability assessments but the correlations are less significant (at the 0.05 level). Figure C.7 gives the results for the replication phase. The only difference is a strong correlation between the navigability estimates and the "site-map assessments".

As introduced, the assessed websites do not have all these navigation elements. Therefore, we focus on the use of these site-level navigation mechanisms.
We analyse if the subjects use the site-level navigation mechanisms for the first phase of the experiment and then for the replication phase. We consider that a subject uses a navigation element (when it is available), if he assesses the usefulness of this navigation element (on a 1 to 4 scale). If he does not use it, or if it is not available, the subject has the possibility to answer "not available or not tested". The use of site-level navigation mechanisms is presented in Table 5.4.

Table 5.4: Observed use of site-level navigation mechanisms.

Nav. Mechanism	% of use (Phase1)	% of use (Replication)
Menu	79%	82%
Search Engine	74%	77%
Site Map	32%	52%

These results show that, when available, most users use the menu (79% and 82%) and the search function (74% and 77%) to explore a website. These results are very similar for the two experiment phases. On the other hand, the site map is rarely used (32% and 52%).
The extensive use of the intern search engine indicates that we have to enhance the classic metaphor, introduced by Pirolli [PC95] that sees users searching for information by identifying and following the most promising links. We need to have a better understanding of how this type of behaviour co-exists with the use of search engines. It might be interesting to try to characterise the search engine instead of simply verifying its existence.
A site map aims to be a plus to traditional navigation techniques. This explains why few people use site maps. When the site map is available, we observe that only 32% of subjects use it during the first experiment phase; and 52% during the replication. This is in agreement with industrial research [TSPN08]. In their report, Todesco and al. explain that "site maps are used rarely". The two main usability guidelines for site maps are "to call it *Site Map* and use this label to consistently link to the site map throughout the site" and "to use a static design". Moreover, we ask tasks on some "small" websites (some with less than 100 pages) and most of the tasks we define in the experiment are "typical" tasks (for example, to find an item on a shop website, to find a form on an administrative website, i.e. each task is linked to the "core business" of the website). Therefore, the

subjects do not need to use a map.

5.2.2 Questionnaire Analysis: Page-level questions

We ask page-level questions in order to describe what has an influence on the subject's perception of navigability when exploring webpages. We remind the page-level questions:
Q1 How well do you estimate the needed time to display a page? (*DisplayTime*)
Q2 How well were you able to locate yourself in the site? (*BeLocate*)
Q3 How well could you identify which links to follow? (*WhereGo*)
Q4 What was your perception of the organisation of navigation elements? (*Orga*)
Q5 How similar were the different pages in the site? (*SimiPages*)
Q6 How similar was the site to others you have visited? (*SimiSites*)
Q7 How easy was it to return to the home page? (*BackHome*)
Q8 Could you find the required information on pages? (*FindInfo*)
Q9 What is your satisfaction of this site? (*Satis*)

In Figures 5.10 and C.8, we analyse the correlations between the navigabilty scores estimated by users and the investigated factors by the page-level questions (*NavMean* and *NavMedian* are the mean and median scores estimated by the subjects about the website navigability (/10). Tables contain the average of estimates (/4) for each of the 10 factors).

Most aspects are positively correlated to the navigability estimates (columns *NavMean* and *NavMedian*). We observe than the most revealing aspects correspond to the subject's ability to locate information or navigation elements in pages.
The three factors that are correlated most strongly with the navigability estimates are "more general" elements: the satisfaction, the pages organisation, and the ease of finding the required information on pages. Moreover, it shows that subjects are logical with themselves as they assess these factors similarly. These three "more general" elements are significantly (at 0.01 level) and strongly correlated with all the other factors (except with the *display time* and the *visited links recognition*).

Five other factors are significantly (at 0.01 level) correlated with all the others

Figure 5.10: Phase 1: Questionnaire analysis - Correlations between the page-level navigability factors.

Experiment phase 1 - Analysis of the questionnaire - page-level questions

Correlations

		NavMEAN	NavMEDIAN	DisplayTime	BeLocate	WhereGo	Orga	SimiPages	SimiSites	BackHome	RecoVisited	FindInfo	Satis
NavMEAN	Pearson Correlation	1	,956**	,296	,595**	,786**	,851**	,638**	,541**	,643**	,302	,890**	,878**
	Sig. (2-tailed)		,000	,112	,001	,000	,000	,000	,002	,000	,105	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
NavMEDIAN	Pearson Correlation	,956**	1	,278	,544**	,755**	,811**	,646**	,490**	,574**	,266	,816**	,822**
	Sig. (2-tailed)	,000		,136	,002	,000	,000	,000	,006	,001	,156	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
DisplayTime	Pearson Correlation	,296	,278	1	,053	,347	,137	-,058	,108	-,027	,013	,317	,254
	Sig. (2-tailed)	,112	,136		,780	,060	,472	,762	,571	,887	,945	,088	,176
	N	30	30	30	30	30	30	30	30	30	30	30	30
BeLocate	Pearson Correlation	,595**	,544**	,053	1	,589**	,795**	,766**	,773**	,675**	,172	,575**	,736**
	Sig. (2-tailed)	,001	,002	,780		,00	,000	,000	,000	,000	,362	,001	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
WhereGo	Pearson Correlation	,786**	,755**	,347	,589**	1	,719**	,474**	,552**	,421*	,431*	,839**	,796**
	Sig. (2-tailed)	,000	,000	,060	,001		,000	,008	,002	,020	,017	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
Orga	Pearson Correlation	,851**	,811**	,137	,795**	,719**	1	,786**	,745**	,774**	,255	,764**	,812**
	Sig. (2-tailed)	,000	,000	,472	,000	,000		,000	,000	,000	,173	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
SimiPages	Pearson Correlation	,638**	,646**	-,058	,766**	,474**	,786**	1	,636**	,792**	,110	,527**	,695**
	Sig. (2-tailed)	,000	,000	,762	,000	,008	,000		,000	,000	,562	,003	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
SimiSites	Pearson Correlation	,541**	,490**	,108	,773**	,552**	,745**	,636**	1	,537**	,396*	,519**	,615**
	Sig. (2-tailed)	,002	,006	,571	,000	,002	,000	,000		,002	,030	,003	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
BackHome	Pearson Correlation	,643**	,574**	-,027	,675**	,421*	,774**	,792**	,537**	1	,157	,539**	,669**
	Sig. (2-tailed)	,000	,001	,887	,000	,020	,000	,000	,002		,408	,002	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
RecoVisited	Pearson Correlation	,302	,266	,013	,172	,431*	,255	,110	,396*	,157	1	,287	,322
	Sig. (2-tailed)	,105	,156	,945	,362	,017	,173	,562	,030	,408		,124	,082
	N	30	30	30	30	30	30	30	30	30	30	30	30
FindInfo	Pearson Correlation	,890**	,816**	,317	,575**	,839**	,764**	,527**	,519**	,539**	,287	1	,883**
	Sig. (2-tailed)	,000	,000	,088	,001	,000	,000	,003	,003	,002	,124		,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
Satis	Pearson Correlation	,878**	,822**	,254	,736**	,796**	,812**	,695**	,615**	,669**	,322	,883**	1
	Sig. (2-tailed)	,000	,000	,176	,000	,000	,000	,000	,000	,000	,082	,000	
	N	30	30	30	30	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

(except with the *display time* and the *visited links recognition*). Three of them are accurately taken into account by our page-level model: the ease of being located, the links quality (*WhereGo* column), the ease of coming back at the home page (*BackHome*). The other two factors are the similarity between pages and the similarity with existing websites. The coherence and similarity of pages on a website is not considered in the quality model. This element can in fact simplify a navigation process, and therefore it might be desirable to include it in our model. However, it might be difficult to assess. Indeed, to assess similarity between webpages, we would need to know what is the design expected by a user. And we would have to assess how much a page deviates from this expectation, what would be difficult to do in a systematic manner.

In contrast, the correlation between the score (columns *NavMean* and *NavMedian*) and the perceived time required to display the pages is low. We assume that most, if not all, subjects were using high-speed connections. Therefore, the display time mentioned is always highly assessed and has no effect on the website evaluation. The assessment of the display time is weakly correlated with all other factors.
Another factor is weakly correlated with the navigability estimates (and with the other factors): it is the recognition of visited links (*RecoVisited* column). We observe that almost all the websites do not take into account the color change of their visited links. A web surfer has to change his browser options if he wants to see this color change. By default, this factor does not affect the subject navigability perception. Therefore, we should perhaps delete or minimize the importance of this page-level metric.

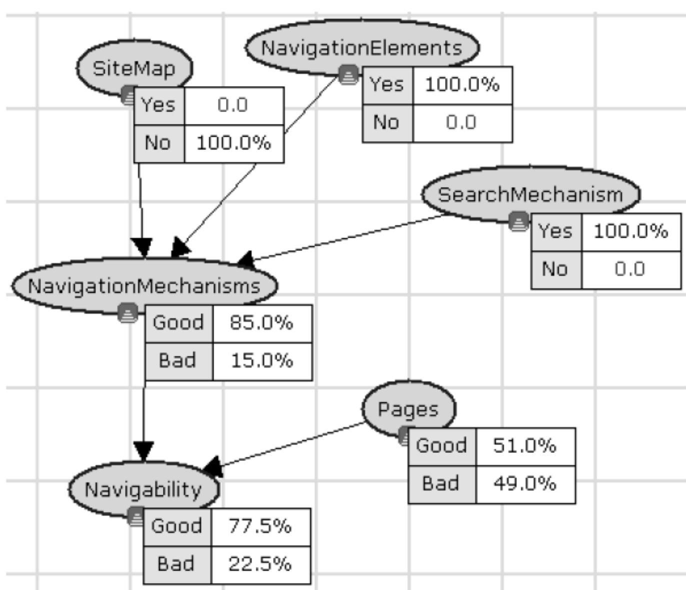
5.3 Improving navigability

In this section we discuss an other possibility offered by our model. Indeed, we use a Bayesian Belief Network (BBN) that allows us to assess the navigability of a website. A BBN is not only used as an assessment tool, but also as an improvement tool.

We look for a successful quality model, that can correctly assess the quality of a website. However, it is also important that our model can improve quality. Based on a previous research[VBSH09], we provide insight as to how we can improve quality. This is important because industries are interested in the relationship between quality evaluation and quality improvement [SGM00].

Figure 5.11 shows, for example, the result of the execution of our model on a good site. The navigability of this site is considered as good with a probability of 77.5%.
The developers/managers could consider that this probability is not high enough and that they could improve the site to increase its navigability. Given the fact that there is a site menu (NavigationElements) and a search engine, there are two possible ways to improve the site:
- improve the navigability of the pages,
- *and/or* add a site map.
The site-wide mechanisms are already thought to be good with P(NavigationMechanisms = good) = 85%, but the page navigability is poor: P(Pages = good) = 51%.

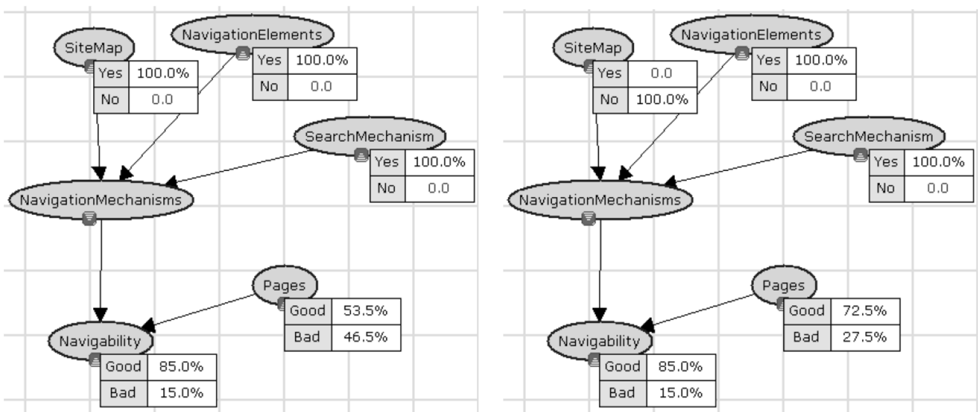
Figure 5.11: Example of an initial quality assessment.



An important property of BBNs is that we can set the probabilities of any node (including the Navigability node) and reevaluate the other probabilities accordingly.

We consider a managerial objective of rising general navigability to a level corresponding to a probability of 85%. A manager could set the output node's value to 85% and try two configurations:
- keep the site without a map (P(SiteMap = Yes) = 0)
- and add a site map (P(SiteMap = Yes) = 1).

Figure 5.12: Example of potential improvements to site: with site map (left); without site map (right).



With a site map, page-level navigability needs to increase to 54% (Figure 5.12, left) to reach the managerial objective of 85% (almost equal to the current value of 51%).
Without the site map, page-level navigability needs to be increased to 72% (Figure 5.12, right).
The development team could then estimate the cost of adding a map and the cost of modifying the pages to find the cheapest solution. For the second solution, as we know the individual navigability of the pages as well as their respective importance, some pages could be targeted. For a particular page, we repeat the output-probability setting to determine the improvement option at page level.

Chapter 6

Improving the Multi-level Model & Future Work

This final chapter aims to suggest ways to improve the "Multi-level Model". It is based on our experiment results presented in Chapter 5. This version of the multi-level model is presented without any concrete implementation and thus without any verification of this model. The second part of this chapter presents the future work. First, it will be obvious that we want to assess the validity our new model. Moreover, we introduce some new research questions, about new supported web technologies, new global navigation elements, or even *Web 3.0* questions.

6.1 Improving the Multi-level Model

After our experiment (cf. chapters 4 & 5) that aims to assess the validity of the multi-level model, we are able to draw some conclusions on this model. In three times, we present a modified version of the "Multi-level Model" that takes our findings into account. First, we remind some key elements of the "Multi-level Model" that we consider as strong points. We explain why we keep these elements in the improved version of the model. Then, we note some modifications that are needed, regarding the most relevant experiment results. Finally, some new elements seem important and should be added to our model. We thus explain what to introduce in the Multi-level Model and how it should be done.

6.1.1 What to keep

Here are the main elements of our Multi-level Model that we want to keep, in order to build an improved version.

Bayesian Network Structure

At the root of our model is a Bayesian Belief Network (BBN) structure, presented in Chapter 3 (cf. section 3.3). We have already discussed the advantages of such an approach (cf. section 3.1.5). A BBN allows us to assess the navigability, from input nodes to a final *navigability node*. However, we show that a BBN also works in reverse, since we can use it to improve website navigability. Therefore, we first assign a certain probability to the final node, and then, we can see to what extent we have to improve certain subnodes (cf. Chapter 5, section 5.3).
Our questionnaire analysis shows that subjects rarely make use of the Site Map when it is available, contrary to their use of the Navigation Menu or the intern Search Function (cf. section 5.2). Therefore, we can justify why our model attaches less importance to the weighting of the Site Map than to the two others global navigation elements (cf. Appendix A; section A.2).

Multi-level Approach

As said previously, the strength of our Multi-level Model lies in its multi-level approach, that distinguishes page- from site-level. This is the main difference with our reference navigability model proposed by Malak (cf. Chapter 3, section 3.2), that is limited to a page-level assessment.
The first priority of the experiment was to assess our multi-level approach. We have seen that the multi-level model could better estimate human judgements than a simple page-level model (cf. section 5.1.3). We can thus confirm the multi-level structure of our model.

6.1.2 What to change

We continue this brief presentation of our "improved Multi-level Model" with some elements that we want to modify.

Probability Review in the BBN

Some changes can be made in the probability tables of our Bayesian Network. These changes should of course be tested and approved by experiments.

At the level of webpages assessment, we have observed that a big majority of subjects does not care about the color change of visited links and breadcrumbs (cf. section 5.2).
In concrete terms, we realize that almost all modern websites do not take into account the color change of their visited links. A web surfer has now to change his browser options [1] if he wants to differenciate unvisited links from visited links. By default, it is thus a criterion that does not affect the user navigability perception. Therefore, we should delete or minimize the importance of this subnode in the Bayesian Network.
Similarly, taking into account the download time of webpages does not to influence our subjects. Consequently, it seems desirable to delete or minimize the importance of this subnode.

Improving the Search Function Assessment

As a result of our questionnaire analysis, we notice that most of our subjects use the intern Search Function when it is available (cf. section 5.2). At present, our navigability model only checks the presence of a Search Function. In view of our observation, we assert that a more accurate score should be computed for the Search Function. Indeed, some search mechanisms could work badly, others could be not visible enough. Thus, we think that our model should take into account the following points:
- the visibility of the search function;
- the presence of the search function at the same place on each webpage;
- (more important) the quality of the results sent back by the search function;
- the presentation of the results.

Simplifying the Aggregation of Pages

A main addition to our Multi-level Model consists of aggregating webpages scores according to different importance algorithms through the composition model (cf. Chapter 3; section 3.3.2). This allows us to compute a global score for webpages, which becomes an input node (in the same way as the 3 global navigation elements) for the calculation of the final navigability node. However, we observe in our experiment results that our different importance strategies are not very dif-

[1] For example, on Mozilla Firefox 5.0, Tools → Options → Content → Colour → *Unauthorize webpages to use their own colours.*

ferent from each other (cf. Chapter 5, section 5.1.4). We conclude that a simple strategy to aggregate webpages scores should be preferred.

6.1.3 What to introduce

We end this section by tackling some new navigation elements that we think interesting to introduce in our model.

Discovering Webpages Templates

We notice in our experiment results that a website navigability score depends more on the number of different **webpages templates** than on the number of webpages. Consequently, we think that it should be interesting to introduce a means of discovering how much templates exist on the website. We believe that in order to properly sample a website, we should include a template identification tool like RoadRunner. Crescenzi et al. [CMM02] propose a system that "works with two HTML pages at a time; pattern discovery is based on the study of similarities and dissimilarities between the pages; mismatches are used to identify relevant structures".

New Navigation Elements

Several new navigation elements could be considered by our model. We can think of different types of navigation menus (e.g. top menu, bottom menu, left- or right-side navigation), login box, etc. For example, at site-level, a new navigation

Figure 6.1: Example of a new navigation element: a tag cloud.

element should be taken into account: it is the tag cloud (as illustrated[2] in Figure 6.1. "A tag cloud (word cloud, or weighted list in visual design) is a visual representation for text data, typically used to depict keyword metadata (tags)

[2]Tag Cloud illustration from `http://www.educause.edu/blog/mpasiewicz/TagCloudofUniversityHomePages/167831`; (Date of access 2011-07-25)

on websites, or to visualize free form text"[3]. Even if this kind of navigation element is especially used in blogs, it could be useful for websites containing a lot of written texts (e.g. newspapers websites).

6.2 Future Work

Now that we have presented several points that we should modify to improve the Multi-level Model, we simply add some other future work that extends our study.

6.2.1 New Model Assessment

First of all, it will be necessary to assess the validity of the model changes we proposed in the previous section. New experiments have to be done in order to check how accurate our new hypotheses are.
For example, here are some areas of research that should be tested:
- To make sure that the improved version of our Multi-level Model is always able to simulate human judgments about the navigability of random websites (preferably more accurately).
- To make sure that the *search engine score* computed by the model is correlated with human judgments about search engine quality.
- To make sure that taking into account new navigation elements (e.g. tag-clouds...) allows to better assess website navigability.
- To make sure that sampling a website by means of a template identification tool is the best way to precisely assess the website navigability.

6.2.2 New Research Questions

We can at last ask some other research questions that broaden the present study. The first one consists of implementing a user interface for our related tool. Second, the Multi-level Model has to assess more types of websites. Then, we briefly tackle the "web 3.0" context with regard to the navigability issue.

Graphic User Interface

A simple improvement of the program should be the implementation of a graphic user interface.

[3]Martin Halvey, An Assessment of Tag Presentation Techniques, on `http://www2007.org/htmlposters/poster988/`; (Date of access 2011-07-25)

More Web Technologies compatible with the model

An important limitation of our model implementation is that our crawler can not collect all types of websites (cf. section 3.4). For example, Flash websites can not be assessed by our program. To improve the model implementation, we should thus add mechanisms that could take into account other web languages.
In this connection, we must also think about "mobile navigation". However, this approach is quite different: the navigability model should be adapted for that. Indeed, users that surf on websites thanks to their smartphones do not have the same navigability issue as the navigability we have studied here. Lots of specific adaptations will be essential from a "smartphone navigation" perspective.

"Semantic Web 3.0" Context

W3C defines semantic web[4] as follows: "In addition to the classic "Web of documents" W3C is helping to build a technology stack to support a "Web of data," the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term "Semantic Web" refers to W3C's vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS."
The "Web 3.0" context puts forward a lot of new questions about future web development. We do not handle this complex problem here. We limit ourselves to introducing the issue. It seems that new unavoidable practices will be asked to future websites developers. Even if the "semantic web" issue does not directly affect the navigation process for users, it will inevitably impact on search results for example. Although the additional web 3.0 layer is intended to be used by computers, web developers should not forget users and will always have to take into account the navigability issue.

[4]From `http://www.w3.org/standards/semanticweb/`; (Date of access 2011-07-14)

Appendix A

A.1 Page-level Model: CPTs of the BBN.

Here are the conditional probability tables (CPTs) of the Bayesian Belief Network (BBN) for the page-level model.

Table A.1: Page-level Model: CPT for the TextFeedBack node

MeaningfulURL	Yes				No			
LinkTitle	Yes		No		Yes		No	
LinkText	Yes	No	Yes	No	Yes	No	Yes	No
Good	0.99	0.3	0.8	0.1	0.9	0.2	0.8	0.01
Bad	0.01	0.7	0.2	0.9	0.1	0.8	0.2	0.99

Table A.2: Page-level Model: CPT for the VisualFeedBack node

CurrentPosLabel	Yes				No			
PathMechanism	Yes		No		Yes		No	
VisitedColour	Yes	No	Yes	No	Yes	No	Yes	No
Good	0.99	0.5	0.6	0.1	0.9	0.4	0.5	0.01
Bad	0.01	0.5	0.4	0.9	0.1	0.6	0.5	0.99

Table A.3: Page-level Model: CPT for the UserFeedBack node

VisualFeedBack	Yes		No	
TextFeedBack	Yes	No	Yes	No
Good	0.99	0.7	0.7	0.4
Bad	0.01	0.3	0.3	0.6

Table A.4: Page-level Model: CPT for the HypertextLinks node

PathMechanism	Yes			No		
NumLinks	High	Medium	Low	High	Medium	Low
Good	0.45	0.5	0.595	0.405	0.45	0.5
Bad	0.55	0.5	0.405	0.595	0.55	0.5

Table A.5: Page-level Model: CPT for the NavigationsOptions node

LinkToHome	Yes		No	
BackButton	Yes	No	Yes	No
Good	0.99	0.7	0.3	0.01
Bad	0.01	0.3	0.7	0.99

Table A.6: Page-level Model: CPT for the Bind node

NavigationOptions	Yes		No	
HyperTextLinks	Yes	No	Yes	No
Good	0.945	0.25	0.75	0.055
Bad	0.055	0.75	0.25	0.945

Table A.7: Page-level Model: CPT for the Navigability node

Bind	Yes				No			
DownSpeed	High		Low		High		Low	
UserFeedBack	Yes	No	Yes	No	Yes	No	Yes	No
Good	0.99	0.55	0.90	0.45	0.55	0.10	0.45	0.01
Bad	0.01	0.45	0.10	0.55	0.45	0.90	0.55	0.99

A.2 Site-level Model: CPTs of the BBN.

Here are the conditional probability tables (CPTs) of the Bayesian Belief Network (BBN) for the site-level model.

Table A.8: Site-level Model: CPT for the NavMechanisms node

SiteMap	Yes				No			
NavMenu	Yes		No		Yes		No	
SearchFunct	Yes	No	Yes	No	Yes	No	Yes	No
Good	0.99	0.7	0.45	0.15	0.85	0.55	0.3	0.01
Bad	0.01	0.3	0.55	0.85	0.15	0.45	0.7	0.99

Table A.9: Site-level Model: CPT for the Navigability node

NavMechanisms	Good		Bad	
Pages	Good	Bad	Good	Bad
Good	0.99	0.7	0.7	0.01
Bad	0.01	0.3	0.3	0.99

Appendix B

B.1 List of the Websites

Table B.1 contains the 30 websites of the experiment (phase 1 & replication).

Table B.1: List of the 30 websites.

Id	Website URL
1	http://www.nature.nps.gov/
2	http://www.estatecafe.com/
3	http://www.edren.org/
4	http://www.environment-agency.gov.uk/
5	http://www.thegrandprixclub.com/
6	http://www.stephengodfreydental.com/
7	http://www.centre-equestre-correze.com/
8	http://www.uciprotour.com/
9	http://www.daghouse.com/
10	http://shop.vans.com/catalog/Vans/en_US/home/index.html
11	http://www.diplomatie.gouv.fr/fr/
12	http://www.weather.com/
13	http://thelogocompany.net/
14	http://www.liaisoncollegeoakville.com/
15	http://www.drakecorp.com/
16	http://www.naturalengland.org.uk/
17	http://www.lewissmith.com/
18	http://www.copyrightservice.co.uk/
19	http://www.cafepress.ca/
20	http://www.triadgolf.com/
21	http://cec-formation.net.pagesperso-orange.fr/index.htm
22	http://www.radioworks.com/hpmain.html
23	http://www.askthemeatman.com/
24	http://www.georgehutchins.com/
25	http://www.theunchartedzone.com/
26	http://www.silver-clay.com/
27	http://www.ravishlondon.com/
28	http://www.orgsites.com/ia/oldtimemusic/
29	http://www.marathon.com/
30	http://www.wregional.com/

B.2 List of the tasks

Here are the experiment tasks (3 different tasks for each website). Altough the questions were asked in French and English in the experiment form, we only write it in English here:

Website 1	`http://www.nature.nps.gov/`
Description	Portal of national parks in the United States.
Task 1:	Find the opening hours for the Canyon Area Visitor Center located near the "Dinosaur National Monument" park, in Colorado.
Task 2:	Find the article of Sally Maertens about her meeting with a volunteer.
Task 3:	Find the article related to the ecosystem restoration.
Answer 1:	`http://www.nps.gov/dino/planyourvisit/hours.htm`
Answer 2:	`http://www.nps.gov/getinvolved/meetavolunteer.htm`
Answer 3:	`http://www.nature.nps.gov/biology/ecosystemrestoration/`

Website 2	`http://www.estatecafe.com/home.html`
Description	Presentation of a coffee mark, La Torcaza Estate.
Task 1:	Find where you can buy this coffee in Europe.
Task 2:	How do you translate "green coffee" in Spanish?
Task 3:	What is the "cupping process" and to which process is it compared?
Answer 1:	`http://www.estatecafe.com/Europe.html`
Answer 2:	`http://www.estatecafe.com/Biblio/greenbeans.html`
Answer 3:	`http://www.estatecafe.com/cupping.html`

Website 3	`http://www.edren.org/`
Description	Information about renal deceases.
Task 1:	Find what should post-transplant patients do in order to travel to countries where a Yellow Fever vaccination certificate is mandatory?
Task 2:	Find the page about diet in renal disease.
Task 3:	Find the board about the historical milestones of kidney transplantation.
Answer 1:	`http://www.edren.org/pages/handbooks/transplant-handbook/vaccinations.php`
Answer 2:	`http://www.edren.org/pages/handbooks/unit-handbook/diet-in-renal-disease.php`
Answer 3:	`http://www.edren.org/pages/history/a-brief-history.php`

Website 4	`http://www.environment-agency.gov.uk/default.aspx`
Description	An environmental agency.
Task 1:	Find where you can download the report about the state of river habitats for the Isle of Man.
Task 2:	Find the detailed map of river levels in the Leicester region (Midlands).
Task 3:	Find information about the registration of a septic tank.
Answer 1:	`http://www.environment-agency.gov.uk/research/library/publications/123383.aspx`
Answer 2:	`http://www.environment-agency.gov.uk/homeandleisure/floods/riverlevels/120549.aspx`
Answer 3:	`http://www.environment-agency.gov.uk/business/topics/water/122003.aspx`

Website 5	`http://www.thegrandprixclub.com/`
Description	Tickets sale for Formula 1 Grand Prix.
Task 1:	Find the reservation form for the Formula 1 Grand Prix of Spa 2011.
Task 2:	Find the offered prices in order to be at the Formula 1 Grand Prix of Montreal 2011.
Task 3:	Find the accurate date of the Formula 1 Grand Prix of Monaco 2011.
Answer 1:	`http://www.thegrandprixclub.com/Formula_1/reservation_form_Belgium.htm`
Answer 2:	`http://www.thegrandprixclub.com/Formula_1/Canada_itinerary.htm`
Answer 3:	`http://www.thegrandprixclub.com/Formula_1/Monaco_2011.htm`

Website 6	http://www.stephengodfreydental.com/index.htm
Description	Regional specialist of Cosmetic Dentistry
Task 1:	Find the time table and precise location of the lectures in Lincoln.
Task 2:	Find the explanation about "Cosmetic Dentistry" (in 2 different pages). Which word can be used as a synonym for "cosmetic"?
Task 3:	Find the price list (year 2007).
Answer 1:	http://www.stephengodfreydental.com/lecturesandtraining.htm
Answer 2:	http://www.stephengodfreydental.com/cosmetic.htm and http://www.stephengodfreydental.com/treatments1.htm
Answer 3:	http://www.stephengodfreydental.com/fees.htm

Website 7	http://www.centre-equestre-correze.com/
Description	A riding school in France.
Task 1:	Find the explanation about the Tipi stays ("Sejours Tipis") offered by this riding school. Then, find the hiring price for a tipi (for one night, one person).
Task 2:	Find the following paragraph: Les grandes lignes de l'education de base enseignee au Petit Canada", about the basic education of this riding school.
Task 3:	Find the document "pinto.doc" giving additional information about the Pinto breeding.
Answer 1:	http://www.centre-equestre-correze.com/tipis.html and http://www.centre-equestre-correze.com/tarifs.html
Answer 2:	http://www.centre-equestre-correze.com/codes.html
Answer 3:	http://www.centre-equestre-correze.com/pinto.html

Website 8	http://www.uciprotour.com/
Description	International cycling union (year 2010).
Task 1:	Find the final ranking of the 2010 Quebec Grand Prix. Then, find the riders list for the RadioShack team in 2010.
Task 2:	Find the total distance of the "Tour des Flandres" on 04/04/2010.
Task 3:	Find the racing cyclist who won the 6th stage of the "Tour de Suisse 2010" (17 Jun 2010: Meiringen - Ponte).
Answer 1:	http://www.uciprotour.com/templates/BUILTIN-NOFRAMES/Template3/layout.asp?MenuId=MTU4MjM&LangId=1
Answer 2:	http://www.uciprotour.com/templates/UCI/UCI5/layout.asp?MenuId=MTU4NDA&LangId=1
Answer 3:	http://www.uciprotour.com/templates/BUILTIN-NOFRAMES/Template3/layout.asp?MenuId=MTU4MjM&LangId=1

Website 9	http://www.daghouse.com/
Description	Music group.
Task 1:	Find the lyrics of the following song "The ambulance song".
Task 2:	Find the creation year of this group, made up of ex-members of Minor Threat and Bloody Mannequin Orchestra.
Task 3:	Find the biography page of Doug Carrion.
Answer 1:	http://www.daghouse.com/lyrics_ambulancesong.html
Answer 2:	http://www.daghouse.com/band.html
Answer 3:	http://www.daghouse.com/bio_carrion.html

Website 10	http://shop.vans.com/catalog/Vans/en_US/home/index.html
Description	Vans shop.
Task 1:	Find the price of the women's shoes "Checkerboard Slip-On".
Task 2:	Find the price of the following item "Solid Ruffled Up Dress".
Task 3:	Find the price of the following item "Vans x Bad Brains Web Belt".
Answer 1:	http://shop.vans.com/catalog/Vans/en_US/product/womens-shoes/classic-shoes/checkerboard-slip-on.html
Answer 2:	http://shop.vans.com/catalog/Vans/en_US/product/womens-clothes/dresses/solid-ruffled-up-dress.html
Answer 3:	http://shop.vans.com/webapp/wcs/stores/servlet/ProductDisplay?productId=669073&storeId=10001&catalogId=10101&langId=-1&vcategoryId=SEARCH

Website 11	http://www.diplomatie.gouv.fr/fr/
Description	French minister of Foreign and European Affairs in France.
Task 1:	Find the following form: a Visa application (type “short stay”) for a foreigner who has the intention of coming to work in France.
Task 2:	Find the place where it is possible to download a .pdf document of the “French directory of international relations” (l’Annuaire francais des relations internationales).
Task 3:	Find the advices given to people that want to stay in a high mountain zone (called risked zone).
Answer 1:	http://www.diplomatie.gouv.fr/fr/france_829/venir-france_4062/entrer-france_4063/colonne-droite_4266/services-formulaires_4269/formulaires-visas_46412.html
Answer 2:	http://www.diplomatie.gouv.fr/fr/actions-france_830/chercheurs-historiens_3119/annuaire-francais-relations-internationales_3123/afri-2008_79592.html
Answer 3:	http://www.diplomatie.gouv.fr/fr/conseils-aux-voyageurs_909/fiches-reflexes_12464/faire_12465/haute-montagne_12478/index.html

Website 12	http://www.weather.com/
Description	Meteorological site from U.S.
Task 1:	Find the 10-days weather forecast for the city of Colombus, Ohio.
Task 2:	Find the weather forecast for this month for the city of Phoenix, Arizona.
Task 3:	Find the weather forecast for the city of Nashville, Tennessee.
Answer 1:	http://www.weather.com/outlook/events/weddings/tenday/USOH0212
Answer 2:	http://www.weather.com/weather/monthly/USAZ0166
Answer 3:	http://www.weather.com/weather/today/USTN0357:1

Website 13	http://thelogocompany.net/
Description	Company for professional logo.
Task 1:	Find the page about ”Yellow Bird Scrapbooking”.
Task 2:	Find the page about ”LionStar Engineering”.
Task 3:	Find the page where you can order “flash website and banner design”. What is the price for a flash banner?
Answer 1:	http://thelogocompany.net/yellow-bird-scrapbooking-case-study.htm
Answer 2:	http://thelogocompany.net/lionstar-engineering-case-study.htm
Answer 3:	https://thelogocompany.net/order-selection.htm

Website 14	http://www.liaisoncollegeoakville.com/index.htm
Description	Cooking school from Oakville, Ontario.
Task 1:	Find the article about “The Eggstraordinary 2007 Student Recipe Contest” and the link to the associated press release (link to download a .pdf document).
Task 2:	Is there an entrance examination for this school?
Task 3:	Find timetables and prices for courses in the campus of Oakville.
Answer 1:	http://www.liaisoncollegeoakville.com/around/eggstraordinary.htm
Answer 2:	http://www.liaisoncollegeoakville.com/about/faq.htm
Answer 3:	http://www.liaisoncollegeoakville.com/recreational_programs/index.htm

Website 15	http://www.drakecorp.com/
Description	Shop of design items.
Task 1:	Find the price of a white chair from the “Elle” collection.
Task 2:	Find the prices for the following item :”Fan Black; folding and stacking chair”.
Task 3:	Find the place where it is possible to download the .pdf brochure about toilet seats.
Answer 1:	http://www.drakecorp.com/p_products/collection.asp?collection=27
Answer 2:	http://www.drakecorp.com/p_products/collection.asp?collection=11
Answer 3:	http://www.drakecorp.com/s_catalogsPDF/2006DR_TOILETSEATS.pdf

Website 16	`http://www.naturalengland.org.uk/`
Description	Natural England is the government's advisor on the natural environment.
Task 1:	Find the article about "Wildlife gardening".
Task 2:	Find the article about "Fossil fuels".
Task 3:	Find the .xls document "List of Regulatory Guidance, Best Practice and Information".
Answer 1:	`http://www.naturalengland.org.uk/advice/wildlifegardening/default.aspx`
Answer 2:	`http://www.naturalengland.org.uk/ourwork/climateandenergy/energy/fossilfuels/default.aspx`
Answer 3:	`http://www.naturalengland.org.uk/ourwork/regulation/wildlife/default.aspx`

Website 17	`http://www.lewissmith.com/`
Description	Enterprise that provides business advices about accountability and taxes.
Task 1:	Find for which taxes categories this enterprise can help its clients.
Task 2:	Find the following article : "Beware the market research call from HMRC" (27/07/2010).
Task 3:	Find the favourite soccer club of Stephen Adderley.
Answer 1:	`http://www.lewissmith.com/SERVICES#`
Answer 2:	`http://www.lewissmith.com/NEWS/July-2010/Beware-the-market-research-call-from-HMRC`
Answer 3:	`http://www.lewissmith.com/PEOPLE`

Website 18	`http://www.copyrightservice.co.uk/`
Description	English site about the copyright.
Task 1:	Find the explanation about the "copyleft" concept.
Task 2:	Find the 10-top myths about the "copyright" topic.
Task 3:	What is the ordering price (per work) of a duplicated copy of your work on line?
Answer 1:	`http://www.copyrightservice.co.uk/copyright/p20_copyleft`
Answer 2:	`http://www.copyrightservice.co.uk/copyright/copyright_myths`
Answer 3:	`http://www.copyrightservice.co.uk/services/price_list`

Website 19	`http://www.cafepress.ca/`
Description	Allow the creation of custom t-shirt.
Task 1:	Find the prices for a humoristic "baby-suit".
Task 2:	Find the price of a wall clock with the Warhol Strat Guitar model.
Task 3:	Find the price of a mug (15oz format) with this model: "Exit, Pursued By A Bear".
Answer 1:	`http://www.cafepress.ca/+humor+baby-bodysuits`
Answer 2:	`http://www.cafepress.ca/+warhol_strat_guitar_wall_clock,34089616`
Answer 3:	`http://www.cafepress.ca/+exits_pursued_by_a_bear_mug,86776143`

Website 20	`http://www.triadgolf.com/`
Description	Triad Golf Today Magazine: website about golf.
Task 1:	Find the phone number of River Landing Golf Course; Greensboro.
Task 2:	Find the article "Mizuno optimizes shaft fitting with new device".
Task 3:	Find the article "Talking Golf with Robert Linville".
Answer 1:	`http://www.triadgolf.com/?page_id=9`
Answer 2:	`http://www.triadgolf.com/?p=370`
Answer 3:	`http://www.triadgolf.com/?p=163`

Website 21	`http://cec-formation.net.pagesperso-orange.fr/index.htm`
Description	Website about medical care. (site de "l'humanitude" dans les soins medicaux - CEC : Communication Etudes Corporelles).
Task 1:	Find the page where you can read the following article "Philosophie des soins; Qu'est ce qu'etre soignant?".
Task 2:	Find the page about the Marie-Line Lamarque thesis : "La non verbalisation de la mort en institution pour personnes agees".
Task 3:	Find the thesis (.pdf file) of Patrice Gaudy called "Depression et suicide de la personne agee : Roles de l'EHPAD".
Answer 1:	`http://cec-formation.net.pagesperso-orange.fr/philo.html`
Answer 2:	`http://cec-formation.net.pagesperso-orange.fr/marieline.html`
Answer 3:	`http://cec-formation.net.pagesperso-orange.fr/memoiregaudy.pdf`

Website 22	`http://www.radioworks.com/hpmain.html`
Description	Radio equipment sale.
Task 1:	Find the price for an "410 Double-male UHF" adaptor.
Task 2:	Find the manual about the following item: "Off-Center-Feed Dipole OCFD_MAX_80" (a .pdf file).
Task 3:	Find the price for an antenna wire (model: "14 Polyethylene Insulated Flex-Weave").
Answer 1:	`http://www.radioworks.com/cconnec2.html`
Answer 2:	`http://www.radioworks.com/OCFD_MAX_80_manual_05_24_10.pdf`
Answer 3:	`http://www.radioworks.com/cwire.html`

Website 23	`http://www.askthemeatman.com/`
Description	Ask The Meatman: meat wholesaler.
Task 1:	Find the page about "How to Plan the Perfect Barbecue" by Debbie Watson.
Task 2:	Find the page about the "Top 10 Cooking Gadgets" by Darcy Miller.
Task 3:	Find the page where it is possible to order an "old time butcher shop black and white poster". What is the price for this poster and the "Beef Poster"?
Answer 1:	`http://www.askthemeatman.com/Plan_the_Perfect_BBQ.htm`
Answer 2:	`http://www.askthemeatman.com/top_10_cooking_gadgets.htm`
Answer 3:	`http://www.askthemeatman.com/old_time_butcher_shop_pork_poster.htm`

Website 24	`http://www.georgehutchins.com/`
Description	Website of George Hutchins for US 2010 congres.
Task 1:	Find the George Hutchins 2010 campaign official 2nd radio commercial.
Task 2:	Find the following document: "Official George Hutchins full-time U.S. regular army discharge document DD214".
Task 3:	Find the page given Hutchins political ideas about "Fight crime - Fight illegal immigration".
Answer 1:	`http://www.georgehutchins.com/index.htm`
Answer 2:	`http://www.georgehutchins.com/hutchins-4-us-congress-8.htm`
Answer 3:	`http://www.georgehutchins.com/hutchins-4-us-congress-4.htm`

Website 25	`http://www.theunchartedzone.com/`
Description	The "uncharted zone" - Musical website.
Task 1:	Find the lyrics of the following song "The Cries in our eyes".
Task 2:	Find the page about "UZ Artist Survey" where it is possible to vote For Your Favorite UZ Artist Once Every 24 Hours.
Task 3:	Find the page about "Music Video Production Rates".
Answer 1:	`http://www.theunchartedzone.com/MarkGormleyTheCriesInOurEyes.htm`
Answer 2:	`http://www.theunchartedzone.com/UZArtistSurvey.htm`
Answer 3:	`http://www.theunchartedzone.com/MusicVideoProduction.htm`

Website 26	`http://www.silver-clay.com/`
Description	Precious metal clay.
Task 1:	Find the price for the following book "Book Art Clay Silver Basic In English".
Task 2:	Find the page where you can see the photo of the following item "River of Gold", created by Brant Palley.
Task 3:	Find the price for the following kiln "Quick Fire 6 KIT! with PCB-1 (Power Control Box) & shelf kit.".
Answer 1:	`http://www.silver-clay.com/nmclay-bin/shop1.pl/page=artclay.htm/SID=1290456106.8291`
Answer 2:	`http://www.silver-clay.com/nmclay-bin/shop1.pl/page=gallery.htm/SID=1290456106.8291`
Answer 3:	`http://www.silver-clay.com/nmclay-bin/shop1.pl/page=Kilns.htm/SID=1290456106.8291`

Website 27	http://www.ravishlondon.com/
Description	London summer festivals.
Task 1:	Find the page about the “Organic Pharmacy” of Soho.
Task 2:	Find the picture called “A pair of Legs” on the page about the London buses.
Task 3:	Find the picture called “Woman Dressed in Green in a Sea of Black & White People” on the page dedicated at the Liverpool Street Station.
Answer 1:	http://www.ravishlondon.com/items/%28954%29.html
Answer 2:	http://www.ravishlondon.com/londonbuses/
Answer 3:	http://www.ravishlondon.com/items/%288%29.html

Website 28	http://www.orgsites.com/ia/oldtimemusic/
Description	Association of National traditional music.
Task 1:	Find the address of the “Pioneer music” museum.
Task 2:	Find when the association took part at the following event: ”American Traditional Music & Dance Fest” in Prague, Czech Republic.
Task 3:	Find the Bob & Sheila’s 2010 play dates.
Answer 1:	http://www.orgsites.com/ia/oldtimemusic/_pgg7.php3
Answer 2:	http://www.orgsites.com/ia/oldtimemusic/_pgg10.php3
Answer 3:	http://www.orgsites.com/ia/oldtimemusic/_pgg4.php3

Website 29	http://www.marathon.com/
Description	Marathon Oil Corporation.
Task 1:	Find the page about ”Shale Development Technology”.
Task 2:	Find the calendar of events.
Task 3:	Find the page about the following product: “Petroleum Coke”.
Answer 1:	http://www.marathon.com/Global_Operations/Technology/Shale_Development_Technology/
Answer 2:	http://phx.corporate-ir.net/phoenix.zhtml?c=84278&p=irol-calendar
Answer 3:	http://www.marathonpetroleum.com/Products/Petroleum_Coke/

Website 30	http://www.wregional.com/
Description	Health care system in Northwest Arkansas.
Task 1:	Find the page about the “Fayetteville City Hospital”. What is the phone number of this hospital?
Task 2:	Find the page about the story of a patient called Donna Hamilton.
Task 3:	Find the page about the nurses program “Medical Surgical Nurse Residency Program”.
Answer 1:	http://www.wregional.com/body.cfm?id=818
Answer 2:	http://www.wregional.com/body.cfm?id=833
Answer 3:	http://www.wregional.com/body.cfm?id=776

B.3 Explanation of the Experiment and Questionnaire

B.3.1 Explanation of the Experiment

Here is the explanation of the experiment that we send to each subject before the experiment.
Website navigability assessment - Antoine Moulart.

Experiment Goal
This experiment aims to assess website navigability.
As an experiment participant, you are going to evaluate 5 different websites.
For each site, it is asked to
- try to perform a task by searching a precise piece of information on the site
- and to answer questions of the linked form.

Task
The task generally consists in searching a precise piece of information (a particular page, article or document) on a website.
Remark regarding the task: During the task, use only the navigation elements of the website. Only the search engine (if there is one) available on the website can be used. Please do not use the possibilities offered by your web browser (as, for example, the back button, the short cut "CTRL + F" (searching function) on Windows, etc.).
In the same idea, please do not use any external search engine (Google for example) to achieve the task. It is indeed important to limit your web session only to the website associated with the task.

Form
Two questions pages are presented for each website. The expected answer time per question should be instantaneous for most questions (check-boxes, score estimation...).
Please answer the questions in the given order, without coming back to previous answers.

Good to know
Your evaluation will be anonymous.
The time you take to perform a task will not be recorded.
The expected time of the entire experiment should be around 30 minutes.
Definitions

Navigability: Navigability is here defined as your ease to access and localise relevant information, thus your facility to solve the asked task. After each task, you will be asked to give your global score for navigation on the basis of this definition.

Navigation Menu: the menu bar containing the large website subdivisions (as a series of tabs or pictures, etc.), generally placed at the top of the page or on the left side.

Site Map, Index: the site map is the site table of contents, containing in a structured way the mean links to the internal subdivisions of the site.

Search Function: the internal search engine included in the website that allows you to type requests or key words in a search field and displays the results.

Explanations about page navigation elements

Ease of knowing "where you are in the website": On some web pages of the site, can you see at which level you are in the website (hierarchical indication, current menu button conspicuously presented, etc.).

Ease of knowing "where to go after" (where will the link lead you?): On a web page, you generally meet some links. Can you know (deduce) to which pages or type of pages the links lead you?

Organisation of navigability elements (position, visibility): How can you judge the organisation quality of the following elements: menu, search function, and links? Are these elements well organised, well situated in order to make your search easy and clear (good visibility)?

Similarity of navigability elements on the different pages: during your navigation session on the site, moving from one page to another, do you find the navigation elements (menu, searching function, etc.) at the same place and presented similarly?

Similarity of navigability elements with other websites that you know well: Is the presentation and organisation of the navigation elements similar to what you usually see on the web?

Ease of coming back to the home page from a random web page on the site: Can you easily come back to the home page with a proposed link; is this link well visible and clear? (Please do not take the "back button" of your web browser into account).

Ease of recognizing visited pages when you leave the page and come back later: On the web pages, can you make the difference between links to the pages that you have already visited and all the other links?

Ease of finding targeted information: Is it usually easy to solve the task (was the task simple? ; were the available navigation elements helpful to perform the task?).

Your satisfaction: Are you personally satisfied after your navigation session on the website?

When you have answered the questions of one task, you can note any remarks about the task or the questions. These remarks will surely be very useful for my future work, so do not hesitate!

Your form:
Online questionnaire URL.

At the end of a form, please click the "submit" button in order to send your answers. The link to the next form will be available in the confirmation message.

B.3.2 Experiment Questionnaire

Here is the English version of the experiment questionnaire.

Web Site Navigability Assessment

Experiment Goal
This experiment aims to assess web site navigability.
As experiment participant, you are going to evaluate 6 different web sites.
For each site, it is asked to
- Try to perform a task by searching a precise piece of information on the site
- and to answer questions of the linked form.

Task
A task consists generally in searching a precise piece of information (a certain page, a certain article or a certain document) on a web site.

Form
For each web site, 2 questions pages are presented. The expected answer time per question should be instantaneous for most of the questions (check-boxes, score estimation...).
Please answer the questions in the given order, without coming back on your previous answers.

Good to know
Your evaluation will be anonymous.
The time you take to perform a task will not be recorded.
The expected time of the entire experiment should be around 30 minutes.

At the end of a form, please click the "submit" button in order to send your answers. The link to the next form will be available in the confirmation message.

Thank you for your participation!

Your Name
Your anonymity will be preserved.

Your mail address

How do you feel when you are browsing on FRENCH web sites ?

	1	2	3	4	5	
It's difficult for me						I'm feeling at ease

How do you feel when you are browsing on ENGLISH web sites ?

	1	2	3	4	5	
It's difficult for me						I'm feeling at ease

Remark regarding the task:
During the task, use only the navigation elements of the web site. Only the search engine (if it exists) available on the web site can be used. Please do not use the possibilities offered by your web browser (like the back button, the short cut "CTRL + F" (searching function) on Windows, etc.).
In the same idea, please do not use any external search engine (like Google for example) to achieve the task. It is thus important to limit your web session only on the web site associated with the task.

Task 1 / 5.
Site : `http://www.uciprotour.com/`
Description : International cycling union web site.
Language : English - French
Task : [EN] Find the total distance of the "Tour des Flandres" on 04/04/2010.
[FR] Trouver la distance totale parcourue lors du Tour des Flandres du 04/04/2010.

Now try to do the task!

1] Did you ever visit this web site ?

Yes	
No	

2] Which score do you give to the navigability of this site ?
The navigability is here defined as your ease to access and localise relevant information, thus your ease to solve the asked task.

	1	2	3	4	5	6	7	8	9	10	
Very bad											Very good

3] Did you manage to perform the task?

Yes	
No	

4] If yes, give your answer ; else can you say why you could not ?
If the asked task was to find a certain page, please copy/paste the founded URL here.

5] How do you estimate the needed time to perform the task?

Instantaneous	
Relatively short	
Relatively long	

6] How do you estimate the following elements:
(From 1: bad to 4: very good)

	Not tested/ doesn't exist	1	2	3	4
Navigation Menu					
Search Function					
Site Map					

7] How do you estimate (the ease of):
(From 1: bad to 4: very good)

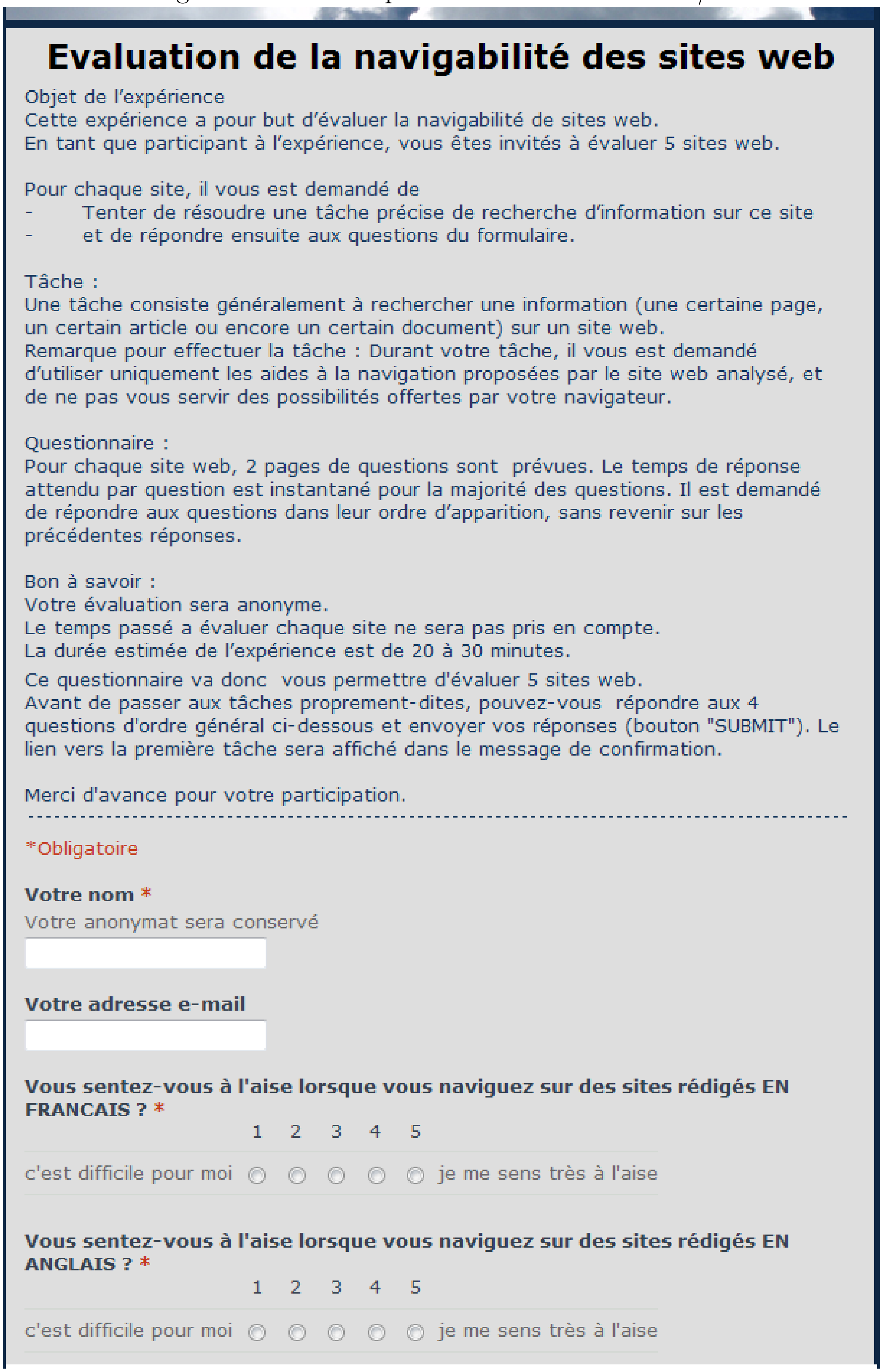
	Not tested	1	2	3	4
Display time of a page					
Knowing where you are in the site					
Knowing "where to go after"					
Organization of navigability elements					
Similarity between the different pages					
Similarity with other web sites					
Coming back at the home page					
Recognizing the visited pages					
Finding the targeted information					
Your satisfaction					

8] Do you have additionnal remarks ?

B.3.3 Screenshots of the Online Questionnaire

Figures B.1, B.2, B.3, and B.4 show the French version of the online questionnaire.

Figure B.1: Online questionnaire - screenshot 1/4.



Evaluation de la navigabilité des sites web

Objet de l'expérience
Cette expérience a pour but d'évaluer la navigabilité de sites web.
En tant que participant à l'expérience, vous êtes invités à évaluer 5 sites web.

Pour chaque site, il vous est demandé de
- Tenter de résoudre une tâche précise de recherche d'information sur ce site
- et de répondre ensuite aux questions du formulaire.

Tâche :
Une tâche consiste généralement à rechercher une information (une certaine page, un certain article ou encore un certain document) sur un site web.
Remarque pour effectuer la tâche : Durant votre tâche, il vous est demandé d'utiliser uniquement les aides à la navigation proposées par le site web analysé, et de ne pas vous servir des possibilités offertes par votre navigateur.

Questionnaire :
Pour chaque site web, 2 pages de questions sont prévues. Le temps de réponse attendu par question est instantané pour la majorité des questions. Il est demandé de répondre aux questions dans leur ordre d'apparition, sans revenir sur les précédentes réponses.

Bon à savoir :
Votre évaluation sera anonyme.
Le temps passé a évaluer chaque site ne sera pas pris en compte.
La durée estimée de l'expérience est de 20 à 30 minutes.

Ce questionnaire va donc vous permettre d'évaluer 5 sites web.
Avant de passer aux tâches proprement-dites, pouvez-vous répondre aux 4 questions d'ordre général ci-dessous et envoyer vos réponses (bouton "SUBMIT"). Le lien vers la première tâche sera affiché dans le message de confirmation.

Merci d'avance pour votre participation.

---

*Obligatoire

Votre nom *
Votre anonymat sera conservé

Votre adresse e-mail

Vous sentez-vous à l'aise lorsque vous naviguez sur des sites rédigés EN FRANCAIS ? *

	1	2	3	4	5	
c'est difficile pour moi	○	○	○	○	○	je me sens très à l'aise

Vous sentez-vous à l'aise lorsque vous naviguez sur des sites rédigés EN ANGLAIS ? *

	1	2	3	4	5	
c'est difficile pour moi	○	○	○	○	○	je me sens très à l'aise

Figure B.2: Online questionnaire - screenshot 2/4.


Evaluation de la navigabilité des si... ×

Evaluation de la navigabilité des sites web

Remarque pour effectuer la tâche :
Il vous est demandé d'utiliser uniquement les aides à la navigation proposées par le site web analysé, et de ne pas vous servir des possibilités offertes par votre navigateur.
De même, veuillez ne pas utiliser de moteur de recherche (externe au site) pour résoudre la tâche. Pour une tâche, limitez donc votre navigation au site associé seulement.

Tache n° 1 sur 5 .

Site : http://shop.vans.com/catalog/Vans/en_US/home/index.html

Brève description : Magasin Vans

Langue du site : Anglais

Tâche : [FR] Trouver le prix de l'article suivant : « Solid Ruffled Up Dress »

[EN] Find the price of the following item « Solid Ruffled Up Dress ».

Essayez maintenant de résoudre la tâche !

* Required

1] Aviez-vous déjà visité ce site auparavant ? *

- Oui
- Non

2] Comment estimez-vous la navigabilité pour ce site? *

La navigabilité est ici définie comme votre facilité à accéder et à localiser de l'information pertinente, donc votre facilité à réussir la tâche demandée.

	1	2	3	4	5	6	7	8	9	10	
Non satisfaisant	○	○	○	○	○	○	○	○	○	○	Très satisfaisant

Figure B.3: Online questionnaire - screenshot 3/4.

3] Avez-vous réussi à accomplir la tâche? *

- Oui
- Non

4] Si oui, donnez votre réponse ; sinon dites pourquoi.

Si la tâche consiste à trouver une certaine page, pouvez-vous copier/coller l'URL de la page en question ? merci.

5] Comment estimez-vous le temps passé pour accomplir la tâche ?

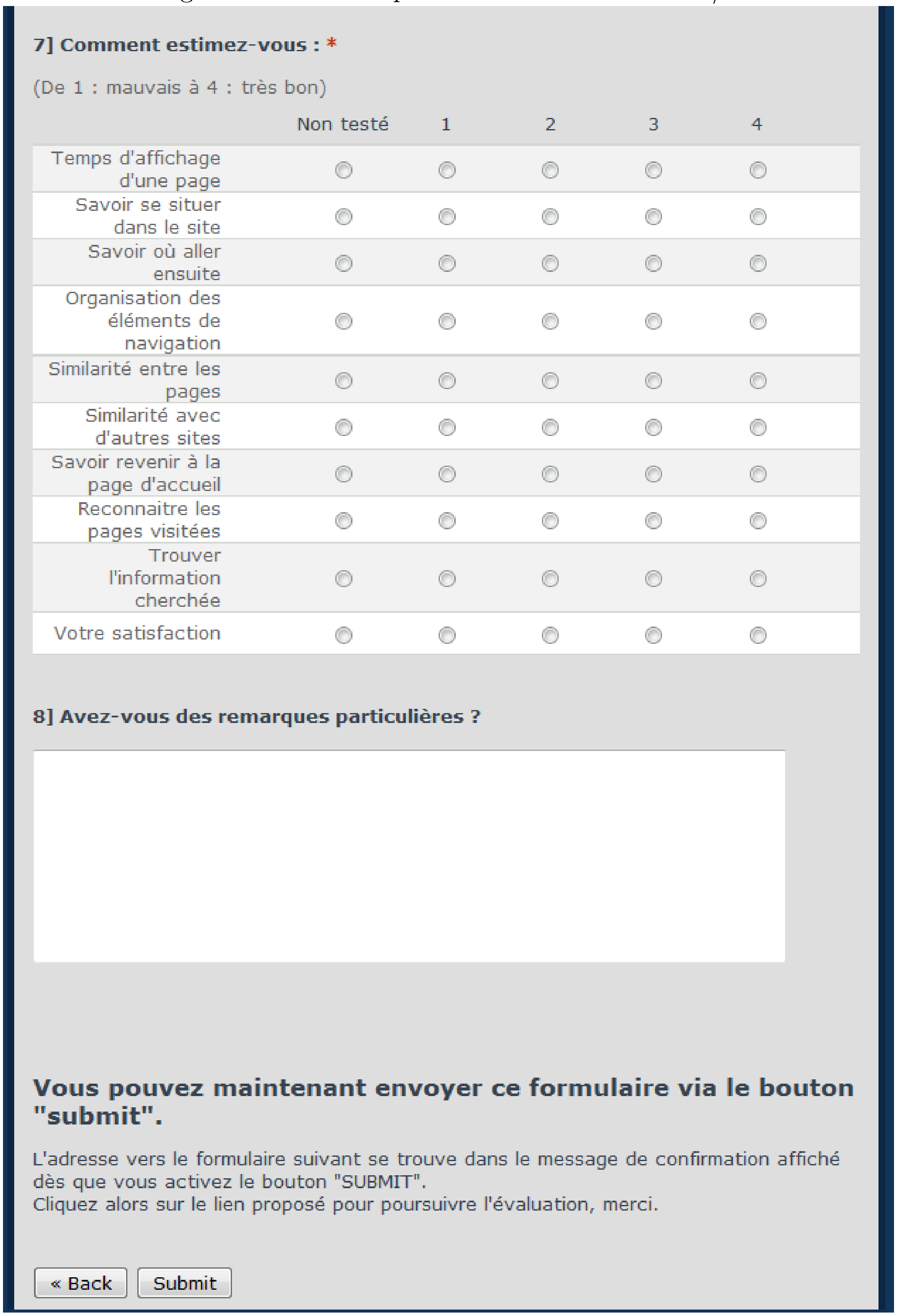
- Instantanné
- Relativement court
- Relativement long

6] Estimez l'utilité des éléments suivants: *

(De 1 : mauvais à 4 : très bon)

	N'existe pas; non testé	1	2	3	4
Menu de navigation					
Fonction de recherche					
Plan du site					

Figure B.4: Online questionnaire - screenshot 4/4.



7] Comment estimez-vous : *

(De 1 : mauvais à 4 : très bon)

	Non testé	1	2	3	4
Temps d'affichage d'une page	○	○	○	○	○
Savoir se situer dans le site	○	○	○	○	○
Savoir où aller ensuite	○	○	○	○	○
Organisation des éléments de navigation	○	○	○	○	○
Similarité entre les pages	○	○	○	○	○
Similarité avec d'autres sites	○	○	○	○	○
Savoir revenir à la page d'accueil	○	○	○	○	○
Reconnaitre les pages visitées	○	○	○	○	○
Trouver l'information cherchée	○	○	○	○	○
Votre satisfaction	○	○	○	○	○

8] Avez-vous des remarques particulières ?

Vous pouvez maintenant envoyer ce formulaire via le bouton "submit".

L'adresse vers le formulaire suivant se trouve dans le message de confirmation affiché dès que vous activez le bouton "SUBMIT".
Cliquez alors sur le lien proposé pour poursuivre l'évaluation, merci.

« Back | Submit

Appendix C

C.1 Program Results

C.1.1 Phase1 - Nav. scores: page- and site-level scores

Phase 1: Table C.1contains the page- and site-level navigability scores.
(assessment on maximum 200 pages; according to the "Visit Probability" algorithm (*pagesVP* and *siteVP*) and the "Simple Mean" strategy (*pagesSM* and *siteSM*) in order to aggregate the pages scores).

Table C.1: Phase 1: Page-level and site-level results - navigability scores on max. 200 pages.

Id	Menu	Search	Map	NbPag	pagesVP	siteVP	pagesSM	siteSM
1	1	1	1	200	0,73509	0,90922	0,73006	0,90774
2	1	0	0	26	0,73452	0,74619	0,73750	0,74756
3	1	1	1	200	0,69724	0,89809	0,67843	0,89256
4	1	1	1	200	0,73636	0,90959	0,73501	0,90919
5	1	0	0	171	0,71275	0,73621	0,71005	0,73498
6	1	0	1	23	0,73783	0,79551	0,72444	0,79002
7	1	0	0	15	0,72669	0,74260	0,72910	0,74371
8	1	1	0	200	0,68459	0,83610	0,69557	0,83995
9	1	0	0	88	0,71080	0,73532	0,71893	0,73904
10	1	1	1	200	0,69322	0,89691	0,69353	0,89700
11	1	1	1	200	0,61253	0,87318	0,61274	0,87325
12	1	1	1	100	0,56687	0,85976	0,57874	0,86325
13	1	0	1	200	0,72684	0,79101	0,71272	0,78521
14	1	0	1	42	0,71438	0,78589	0,72072	0,78850
15	1	0	1	108	0,74000	0,79640	0,77152	0,80932
16	1	1	1	200	0,69270	0,89676	0,69693	0,89800
17	1	0	0	84	0,77915	0,76665	0,78356	0,76867
18	1	0	1	166	0,68367	0,77330	0,66359	0,76507
19	1	1	1	200	0,62185	0,87592	0,64340	0,88226
20	1	1	0	193	0,72630	0,85070	0,74051	0,85568
21	0	1	0	104	0,61489	0,56749	0,62092	0,57092
22	0	0	0	111	0,71098	0,50463	0,71071	0,50445
23	0	1	1	200	0,46790	0,55913	0,49174	0,57128
24	0	0	0	16	0,61449	0,43844	0,62296	0,44425
25	0	1	0	59	0,77469	0,65857	0,77479	0,65863
26	1	0	0	200	0,64813	0,70659	0,64773	0,70641
27	0	0	0	12	0,47788	0,34473	0,57490	0,41128
28	1	0	0	10	0,65424	0,70939	0,59699	0,68315
29	1	1	1	151	0,72390	0,90593	0,71021	0,90190
30	1	1	1	154	0,69369	0,89705	0,65286	0,88504

C.1.2 Phase1 - Nav. scores - site-level - max. 50 pages

Table C.2 contains the navigability scores (computed on maximum 50 pages) for the 30 websites.

Table C.2: Phase 1: Site-level results: navigability scores on max. 50 pages.

id	nb Pg	Visit Proba	Simple Mean	Weight NIpath	Betw. Centr.	Rand. Walk.	Page Rank	HITS (hub)	Mark Centr.	K-Step Markov
1	50	0,9132	0,9128	0,9122	0,9197	0,9152	0,9130	0,9128	0,9130	0,9142
2	24	0,7471	0,7479	0,7469	0,7405	0,7455	0,7478	0,7479	0,7476	0,7458
3	50	0,9047	0,9072	0,9086	0,9139	0,9102	0,9068	0,9072	0,9069	0,9122
4	50	0,9049	0,9064	0,9067	0,9064	0,9064	0,9063	0,9064	0,9063	0,9067
5	26	0,7270	0,7249	0,7245	0,7021	0,7230	0,7253	0,7249	0,7252	0,7228
6	20	0,7978	0,7900	0,7910	0,7867	0,7969	0,7903	0,7900	0,7902	0,8009
7	15	0,7426	0,7437	0,7436	0,6387	0,7430	0,7437	0,7437	0,7437	0,7426
8	50	0,8463	0,8503	0,8507	0,8528	0,8509	0,8498	0,8503	0,8499	0,8503
9	50	0,7348	0,7351	0,7322	0,7084	0,7287	0,7350	0,7351	0,7346	0,7297
10	50	0,9004	0,8986	0,8994	0,8859	0,8976	0,8986	0,8986	0,8986	0,9021
11	50	0,8714	0,8718	0,8718	0,8401	0,8727	0,8718	0,8718	0,8718	0,8731
12	50	0,8410	0,8637	0,8628	0,8401	0,8487	0,8637	0,8637	0,8637	0,8410
13	47	0,7850	0,7842	0,7833	0,7808	0,7798	0,7846	0,7842	0,7846	0,7769
14	21	0,7861	0,7852	0,7845	0,7900	0,7843	0,7856	0,7852	0,7856	0,7826
15	50	0,7817	0,8025	0,8000	0,8150	0,7847	0,8027	0,8025	0,8027	0,7678
16	50	0,9011	0,9029	0,9028	0,8401	0,9017	0,9029	0,9029	0,9029	0,9011
17	17	0,7529	0,7491	0,7441	0,7686	0,7567	0,7488	0,7491	0,7474	0,7465
18	50	0,7702	0,7623	0,7626	0,6980	0,7675	0,7623	0,7623	0,7623	0,7702
19	50	0,8756	0,8909	0,8902	0,8401	0,8795	0,8909	0,8909	0,8909	0,8737
20	38	0,8483	0,8540	0,8537	0,7715	0,8502	0,8540	0,8540	0,8540	0,8483
21	50	0,5334	0,5734	0,5718	0,5020	0,5469	0,5734	0,5734	0,5734	0,5334
22	50	0,5075	0,5054	0,5063	0,5140	0,5060	0,5057	0,5054	0,5058	0,5049
23	50	0,5418	0,5512	0,5508	0,5755	0,5450	0,5512	0,5512	0,5512	0,5418
24	16	0,4398	0,4450	0,4409	0,4740	0,4364	0,4465	0,4450	0,4466	0,4243
25	45	0,6657	0,6648	0,6641	0,6667	0,6651	0,6650	0,6648	0,6650	0,6637
26	50	0,7047	0,7050	0,7052	0,7083	0,7069	0,7048	0,7050	0,7048	0,7082
27	12	0,3447	0,4113	0,4000	0,3599	0,3682	0,4123	0,4113	0,4128	0,3447
28	10	0,7113	0,6835	0,6888	0,6387	0,7017	0,6829	0,6835	0,6827	0,7113
29	50	0,9015	0,9016	0,9016	0,8401	0,9015	0,9016	0,9016	0,9016	0,9015
30	50	0,8857	0,8830	0,8831	0,8401	0,8848	0,8830	0,8830	0,8830	0,8857

C.1.3 Phase1 - Nav. scores - site-level - max. 200 pages

Table C.3 contains the navigability scores (computed on maximum 200 pages) for the 30 websites.

Table C.3: Phase 1: Site-level results: navigability scores on max. 200 pages.

id	nb Pg	Visit Proba.	Simple Mean	Weight NIpath	Betw. Centr.	Rand. Walk.	Page Rank	HITS (hub)	Mark Centr.	K-Step Markov
1	200	0,9092	0,9077	0,9086	0,9135	0,9116	0,9078	0,9077	0,9078	0,9140
2	26	0,7462	0,7476	0,7463	0,7405	0,7443	0,7475	0,7476	0,7474	0,7437
3	200	0,8981	0,8926	0,8948	0,8980	0,8986	0,8926	0,8926	0,8927	0,9109
4	200	0,9096	0,9092	0,9087	0,8834	0,9002	0,9092	0,9092	0,9092	0,9051
5	171	0,7362	0,7350	0,7330	0,7296	0,7303	0,7356	0,7350	0,7356	0,7271
6	23	0,7955	0,7900	0,7910	0,7908	0,7967	0,7903	0,7900	0,7903	0,8004
7	15	0,7426	0,7437	0,7436	0,6387	0,7430	0,7437	0,7437	0,7437	0,7426
8	200	0,8361	0,8399	0,8437	0,8476	0,8458	0,8401	0,8399	0,8402	0,8497
9	88	0,7353	0,7390	0,7366	0,7034	0,7287	0,7389	0,7390	0,7386	0,7298
10	200	0,8969	0,8970	0,8973	0,8993	0,8992	0,8970	0,8970	0,8970	0,9005
11	200	0,8732	0,8732	0,8721	0,8335	0,8612	0,8733	0,8732	0,8733	0,8632
12	100	0,8598	0,8632	0,8627	0,8401	0,8487	0,8633	0,8632	0,8633	0,8406
13	200	0,7910	0,7852	0,7847	0,7703	0,7773	0,7855	0,7852	0,7855	0,7782
14	42	0,7859	0,7885	0,7867	0,7890	0,7860	0,7884	0,7885	0,7883	0,7806
15	108	0,7964	0,8093	0,8056	0,7655	0,7800	0,8094	0,8093	0,8093	0,7697
16	200	0,8968	0,8980	0,8994	0,9021	0,9001	0,8979	0,8980	0,8979	0,9018
17	84	0,7667	0,7687	0,7604	0,7411	0,7569	0,7682	0,7687	0,7675	0,7461
18	166	0,7733	0,7651	0,7647	0,7759	0,7707	0,7656	0,7651	0,7656	0,7711
19	200	0,8759	0,8823	0,8840	0,8595	0,8708	0,8823	0,8823	0,8823	0,8733
20	193	0,8507	0,8557	0,8549	0,8517	0,8533	0,8522	0,8557	0,8517	0,8480
21	104	0,5675	0,5709	0,5699	0,5894	0,5542	0,5700	0,5709	0,5700	0,5358
22	111	0,5046	0,5044	0,5068	0,5122	0,5069	0,5042	0,5044	0,5042	0,5050
23	200	0,5591	0,5713	0,5661	0,5108	0,5503	0,5713	0,5713	0,5712	0,5450
24	16	0,4384	0,4442	0,4401	0,4724	0,4352	0,4457	0,4442	0,4458	0,4230
25	59	0,6586	0,6586	0,6611	0,6623	0,6613	0,6590	0,6586	0,6591	0,6627
26	200	0,7066	0,7064	0,7063	0,6923	0,7025	0,7064	0,7064	0,7064	0,7082
27	12	0,3447	0,4113	0,4000	0,3599	0,3682	0,4123	0,4113	0,4128	0,3447
28	10	0,7094	0,6831	0,6885	0,6387	0,7000	0,6826	0,6831	0,6822	0,7094
29	151	0,9059	0,9019	0,9019	0,8639	0,8923	0,9022	0,9019	0,9022	0,9020
30	154	0,8970	0,8850	0,8842	0,8904	0,8873	0,8858	0,8850	0,8858	0,8860

C.1.4 Replication - Nav. scores - site-level - max. 50 pages

Table C.4 contains the navigability scores (computed on maximum 50 pages) for the 30 websites.

Table C.4: Replication; Site-level results: navigability scores on max. 50 pages.

id	nb Pg	Visit Proba.	Simple Mean	Weight NIpath	Betw. Centr.	Rand. Walk.	Page Rank	HITS (hub)	Mark Centr.	K-Step Markov
1	50	0,9046	0,9111	0,9114	0,9194	0,9144	0,9104	0,9111	0,9104	0,9133
2	26	0,7462	0,7476	0,7463	0,7405	0,7443	0,7475	0,7476	0,7474	0,7437
3	50	0,9048	0,9072	0,9085	0,9140	0,9103	0,9068	0,9072	0,9069	0,9122
4	50	0,9026	0,9047	0,9049	0,9064	0,9042	0,9046	0,9047	0,9047	0,9039
5	50	0,7349	0,7328	0,7298	0,7221	0,7273	0,7334	0,7328	0,7332	0,7243
6	23	0,7955	0,7901	0,7910	0,7932	0,7971	0,7904	0,7901	0,7903	0,8005
7	15	0,7426	0,7437	0,7436	0,6387	0,7430	0,7437	0,7437	0,7437	0,7426
8	50	0,8509	0,8518	0,8519	0,8532	0,8515	0,8518	0,8518	0,8518	0,8507
9	50	0,7347	0,7351	0,7321	0,7084	0,7287	0,7350	0,7351	0,7346	0,7296
10	50	0,8992	0,8968	0,8972	0,8806	0,8946	0,8968	0,8968	0,8968	0,9006
11	50	0,8570	0,8612	0,8610	0,8401	0,8584	0,8612	0,8612	0,8612	0,8570
12	50	0,8437	0,8600	0,8594	0,8401	0,8492	0,8600	0,8600	0,8600	0,8437
13	47	0,7848	0,7833	0,7825	0,7857	0,7802	0,7837	0,7833	0,7837	0,7766
14	21	0,8032	0,8030	0,8031	0,6980	0,8031	0,8030	0,8030	0,8030	0,8032
15	50	0,7816	0,8024	0,8001	0,8146	0,7842	0,8026	0,8024	0,8026	0,7679
16	50	0,9025	0,9021	0,9022	0,8401	0,9024	0,9021	0,9021	0,9021	0,9025
17	19	0,7523	0,7509	0,7451	0,7672	0,7574	0,7504	0,7509	0,7488	0,7456
18	50	0,7697	0,7620	0,7623	0,6980	0,7671	0,7619	0,7620	0,7619	0,7697
19	50	0,8670	0,8813	0,8807	0,8401	0,8718	0,8813	0,8813	0,8813	0,8670
20	44	0,8501	0,8544	0,8543	0,8630	0,8518	0,8544	0,8544	0,8544	0,8488
21	50	0,5334	0,5734	0,5718	0,5020	0,5469	0,5734	0,5734	0,5734	0,5334
22	50	0,5075	0,5023	0,5028	0,5205	0,5066	0,5028	0,5023	0,5028	0,5042
23	50	0,5442	0,5540	0,5536	0,5755	0,5475	0,5540	0,5540	0,5540	0,5442
24	16	0,4403	0,4453	0,4396	0,4599	0,4327	0,4472	0,4453	0,4472	0,4219
25	45	0,6606	0,6620	0,6625	0,6531	0,6607	0,6616	0,6620	0,6615	0,6633
26	50	0,7068	0,7059	0,7056	0,7062	0,7072	0,7058	0,7059	0,7058	0,7087
27	12	0,5135	0,5135	0,5135	0,3599	0,3599	0,5135	0,5135	0,5135	0,5135
28	10	0,7113	0,6835	0,6888	0,6387	0,7017	0,6829	0,6835	0,6827	0,7113
29	50	0,9028	0,9025	0,9025	0,8401	0,9027	0,9025	0,9025	0,9025	0,9028
30	50	0,8846	0,8817	0,8818	0,8401	0,8836	0,8817	0,8817	0,8817	0,8846

C.2 Subjects Assessments

Table C.5 contains the navigability score computed according to our Multi-level Model (by means of the Visit Probability strategy) and the subjects assessments (/10; mean and median scores) for the experiment first phase and for the replication.

Table C.5: Phase 1 & Replication : Site-level scores (according to the Visit Proba. algorithm); subjects assessments (mean and median scores).

Id	Phase1 scoreVP	Subj. Mean	Subj. Median	Repli. scoreVP	Subj. Mean	Subj. Median
1	0,90922	7,50	7,00	0,90461	6,33	7
2	0,74619	4,33	5,00	0,74619	3,33	3
3	0,89809	4,67	5,00	0,90479	7,00	8
4	0,90959	7,67	8,00	0,90261	5,33	6
5	0,73621	6,50	5,50	0,73487	7,00	7
6	0,79551	7,25	8,50	0,79552	5,67	5
7	0,74260	7,33	10,00	0,74260	9,00	9
8	0,83610	8,00	8,00	-	-	-
9	0,73532	4,33	3,00	0,73473	9,00	9
10	0,89691	8,33	10,00	0,89919	9,33	9
11	0,87318	5,33	5,00	0,85703	6,00	6
12	0,85976	8,33	9,00	0,84373	8,67	9
13	0,79101	4,00	2,50	0,78481	7,33	8
14	0,78589	5,00	4,50	0,80315	6,33	8
15	0,79640	8,00	8,00	0,78158	6,33	7
16	0,89676	5,33	6,00	0,90252	8,33	8
17	0,76665	8,00	9,00	0,75233	8,00	9
18	0,77330	4,67	4,00	0,76972	9,33	9
19	0,87592	8,33	9,00	0,86701	6,33	6
20	0,85070	7,00	8,00	0,85014	5,33	4
21	0,56749	2,50	2,50	0,53341	2,00	2
22	0,50463	4,75	4,50	0,50754	1,00	1
23	0,55913	2,75	2,50	0,54421	2,00	2
24	0,43844	1,40	1,00	0,44034	3,00	1
25	0,65857	3,00	3,50	0,66057	3,00	3
26	0,70659	3,25	2,00	0,70680	4,33	4
27	0,34473	3,50	2,50	0,51349	1,00	1
28	0,70939	4,00	4,00	0,71129	4,33	3
29	0,90593	7,75	8,00	-	-	-
30	0,89705	7,80	8,00	0,88459	8,67	9

C.3 Analysis of the Replication of the Experiment

C.3.1 Page-level analysis

We present the results of the page-level analysis for the replication phase. The results interpretation is explained in Chapter 5 (cf. section 5.1.2).

Figure C.1: Page-level analysis: Normality assumption: Kolmogorov-Smirnov test.

One-Sample Kolmogorov-Smirnov Test

		scorePages SM	subjMean Score	subjMedian Score
N		28	28	28
Normal Parameters[a,b]	Mean	,6836270461	5,83214	5,82
	Std. Deviation	...	2,589104	2,881
Most Extreme Differences	Absolute	,225	,112	,168
	Positive	,122	,088	,135
	Negative	-,225	-,112	-,168
Kolmogorov-Smirnov Z		1,192	,592	,889
Asymp. Sig. (2-tailed)		,117	,874	,407

a. Test distribution is Normal.

b. Calculated from data.

Figure C.2: Page-level analysis: correlations between the page-level scores (computed by the composition model) and human judgments about navigability.

Correlations

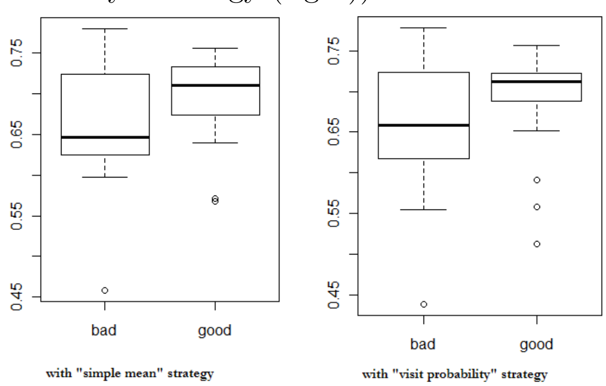
		scorePages VP	scorePages SM	subjMean Score	subjMedian Score
scorePagesVP	Pearson Correlation	1	,949**	,164	,198
	Sig. (2-tailed)		,000	,405	,313
	N	28	28	28	28
scorePagesSM	Pearson Correlation	,949**	1	,150	,214
	Sig. (2-tailed)	,000		,445	,275
	N	28	28	28	28
subjMeanScore	Pearson Correlation	,164	,150	1	,967**
	Sig. (2-tailed)	,405	,445		,000
	N	28	28	28	28
subjMedianScore	Pearson Correlation	,198	,214	,967**	1
	Sig. (2-tailed)	,313	,275	,000	
	N	28	28	28	28

**. Correlation is significant at the 0.01 level (2-tailed).

Table C.6: Replication: Page-level analysis: Population test.

t-test param.	t	df	p-value	Mean (good)	Mean (bad)
scorePagesSM by SubjMeanScore > 5	-1.1905	10.671	0.2597	0.697	0.655
scorePagesSM by SubjMedianScore > 5	-0.7607	15.291	0.4584	0.693	0.669
scorePagesVP by SubjMeanScore > 5	-0.9535	11.067	0.3607	0.686	0.652
scorePagesVP by SubjMedianScore > 5	-0.4953	16.021	0.6271	0.683	0.666

Figure C.3: Page-level analysis: Compared distribution of scores for good and bad sites (navigability scores computed by means of a composition model (that aggregates the page-level scores according to the "Simple Mean" strategy (left) and the "Visit Probability" strategy (right)).



C.3.2 Site-level analysis

We present the results of the site-level analysis for the replication phase. The results interpretation is explained in Chapter 5 (cf. section 5.1.3).

C.3.3 Questionnaire Analysis

The results interpretation of the tables C.7 and C.8 is explained in Chapter 5 (cf. section 5.2).

Figure C.8 presents the correlations between the navigabilty scores estimated by users and the investigated factors by the page-level questions (NavMean and NavMedian are the mean and median score estimated by the subjects about the website navigability (/10). The table contains the average of estimates (/4) for each of the 10 factors).

C.3.4 Impact of the Weighting Algorithms

The results interpretation of the Table C.9 is explained in Chapter 5 (cf. section 5.1.4).

Figure C.4: Site-level analysis: Normality assumption: Kolmogorov-Smirnov test.

One-Sample Kolmogorov-Smirnov Test

		scoreSiteVP	scoreSiteSM	subjMean Score	subjMedian Score
N		28	28	28	28
Normal Parameters[a,,b]	Mean	,7528341507	,7554087457	5,83214	5,82
	Std. Deviation	...	...	2,589104	2,881
Most Extreme Differences	Absolute	,161	,148	,112	,168
	Positive	,132	,124	,088	,135
	Negative	-,161	-,148	-,112	-,168
Kolmogorov-Smirnov Z		,853	,782	,592	,889
Asymp. Sig. (2-tailed)		,460	,573	,874	,407

a. Test distribution is Normal.

b. Calculated from data.

Figure C.5: Site-level analysis: correlations between websites navigability scores (computed by the multi-level model) and human judgments.

Correlations

		scoreSiteVP	scoreSiteSM	subjMean Score	subjMedian Score
scoreSiteVP	Pearson Correlation	1	,997**	,740**	,755**
	Sig. (2-tailed)		,000	,000	,000
	N	28	28	28	28
scoreSiteSM	Pearson Correlation	,997**	1	,732**	,753**
	Sig. (2-tailed)	,000		,000	,000
	N	28	28	28	28
subjMeanScore	Pearson Correlation	,740**	,732**	1	,967**
	Sig. (2-tailed)	,000	,000		,000
	N	28	28	28	28
subjMedianScore	Pearson Correlation	,755**	,753**	,967**	1
	Sig. (2-tailed)	,000	,000	,000	
	N	28	28	28	28

**. Correlation is significant at the 0.01 level (2-tailed).

Table C.7: Replication: Site-level analysis : Population test.

t-test param.	t	df	p-value	Mean (good)	Mean (bad)
scoreSiteSM by SubjMeanScore > 5	-6.0828	11.092	7.653e-05	0.830	0.599
scoreSiteSM by SubjMedianScore > 5	-4.464	13.605	0.0006	0.830	0.639
scoreSiteVP by SubjMeanScore > 5	-5.8275	10.623	0.0001	0.827	0.596
scoreSiteVP by SubjMedianScore > 5	-4.3307	13.227	0.0008	0.827	0.637

Figure C.6: Site-level analysis: Compared distribution of scores for good and bad sites (navigability scores computed by means of the multi-level model (pages scores aggregated according to the "Simple Mean" strategy (left) and the "Visit Probability" strategy (right)).

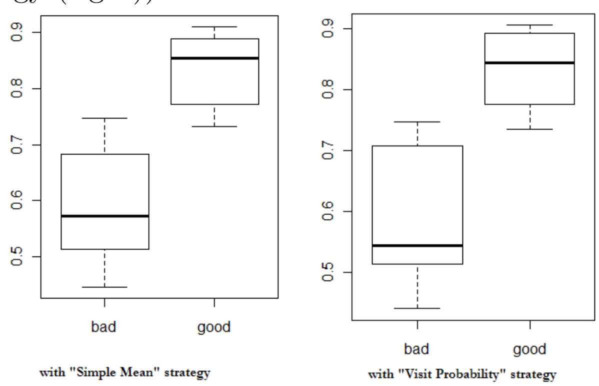


Figure C.7: Replication: Questionnaire analysis - Correlations between the estimates about the site-level navigation elements and the navigability estimates.

Replication

Correlations

		menu	search	siteMap
mean	Pearson Correlation	,734**	,470*	,817**
	Sig. (2-tailed)	,000	,037	,000
median	Pearson Correlation	,695**	,516*	,743**
	Sig. (2-tailed)	,000	,020	,000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure C.8: Replication: Questionnaire analysis - Correlations between the page-level navigability factors estimated by subjects.

Experiment replication phase: Analysis of the questionnaire - page-level questions

Correlations

		NavMEAN	NavMEDIAN	DisplayTime	BeSituated	WhereGo	Organisation	Similarity Pages	SimilaritySites	BackHome	RecoVisited	FindInfo	Satisfaction
NavMEAN	Pearson Correlation	1	,967**	,313	,675**	,821**	,858**	,770**	,409*	,567**	,089	,826**	,850**
	Sig. (2-tailed)		,000	,092	,000	,000	,000	,000	,025	,001	,640	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
NavMEDIAN	Pearson Correlation	,967**	1	,191	,622**	,771**	,798**	,757**	,390*	,550**	,136	,787**	,792**
	Sig. (2-tailed)	,000		,311	,000	,000	,000	,000	,033	,002	,473	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
DisplayTime	Pearson Correlation	,313	,191	1	,324	,323	,322	,187	-,097	,365*	-,150	,168	,317
	Sig. (2-tailed)	,092	,311		,080	,082	,083	,321	,610	,047	,429	,375	,088
	N	30	30	30	30	30	30	30	30	30	30	30	30
BeSituated	Pearson Correlation	,675**	,622**	,324	1	,724**	,748**	,802**	,384*	,586**	,204	,603**	,792**
	Sig. (2-tailed)	,000	,000	,080		,000	,000	,000	,036	,001	,281	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
WhereGo	Pearson Correlation	,821**	,771**	,323	,724**	1	,818**	,759**	,387*	,608**	,220	,768**	,794**
	Sig. (2-tailed)	,000	,000	,082	,000		,000	,000	,035	,000	,243	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
Organisation	Pearson Correlation	,858**	,798**	,322	,748**	,818**	1	,685**	,457*	,671**	,246	,779**	,937**
	Sig. (2-tailed)	,000	,000	,083	,000	,000		,000	,011	,000	,189	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
SimilarityPages	Pearson Correlation	,770**	,757**	,187	,802**	,759**	,685**	1	,442*	,666**	,134	,611**	,736**
	Sig. (2-tailed)	,000	,000	,321	,000	,000	,000		,014	,000	,480	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
SimilaritySites	Pearson Correlation	,409*	,390*	-,097	,384*	,387*	,457*	,442*	1	,421*	,203	,349	,404*
	Sig. (2-tailed)	,025	,033	,610	,036	,035	,011	,014		,020	,281	,059	,027
	N	30	30	30	30	30	30	30	30	30	30	30	30
BackHome	Pearson Correlation	,567**	,550**	,365*	,586**	,608**	,671**	,666**	,421*	1	,273	,536**	,675**
	Sig. (2-tailed)	,001	,002	,047	,001	,000	,000	,000	,020		,144	,002	,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
RecoVisited	Pearson Correlation	,089	,136	-,150	,204	,220	,246	,134	,203	,273	1	,198	,275
	Sig. (2-tailed)	,640	,473	,429	,281	,243	,189	,480	,281	,144		,295	,142
	N	30	30	30	30	30	30	30	30	30	30	30	30
FindInfo	Pearson Correlation	,826**	,787**	,168	,603**	,768**	,779**	,611**	,349	,536**	,198	1	,839**
	Sig. (2-tailed)	,000	,000	,375	,000	,000	,000	,000	,059	,002	,295		,000
	N	30	30	30	30	30	30	30	30	30	30	30	30
Satisfaction	Pearson Correlation	,850**	,792**	,317	,792**	,794**	,937**	,736**	,404*	,675**	,275	,839**	1
	Sig. (2-tailed)	,000	,000	,088	,000	,000	,000	,000	,027	,000	,142	,000	
	N	30	30	30	30	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure C.9: Correlations between navigability scores of the 30 websites (according to different weighting algorithms to aggregate webpages scores), and with mean and median navigability scores estimated by subjects.

Correlations (Replication)

		Visit Probability	SimpleMeanOf PagesScores	Weighted_NI _Paths	Betweenness Centrality	RandomWalkB etweenness	PageRanks	HITS_hub	Markov Centrality	Kstep Markov	MEAN_subjects	MEDIAN_subjects
VisitProbability	Pearson Correlation	1	,997**	,997**	,951**	,984**	,997**	,997**	,997**	,999**	,677**	,671**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
SimpleMeanOf PagesScores	Pearson Correlation	,997**	1	1,000**	,955**	,984**	1,000**	1,000**	1,000**	,996**	,670**	,670**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Weighted_NI_ Paths	Pearson Correlation	,997**	1,000**	1	,954**	,984**	1,000**	1,000**	1,000**	,997**	,666**	,666**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Betweenness Centrality	Pearson Correlation	,951**	,955**	,954**	1	,969**	,955**	,955**	,955**	,948**	,610**	,604**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
RandomWalk Betweenness	Pearson Correlation	,984**	,984**	,984**	,969**	1	,984**	,984**	,984**	,982**	,682**	,672**
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
PageRanks	Pearson Correlation	,997**	1,000**	1,000**	,955**	,984**	1	1,000**	1,000**	,996**	,670**	,670**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
HITS_hub	Pearson Correlation	,997**	1,000**	1,000**	,955**	,984**	1,000**	1	1,000**	,996**	,670**	,670**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
Markov Centrality	Pearson Correlation	,997**	1,000**	1,000**	,955**	,984**	1,000**	1,000**	1	,996**	,670**	,669**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000		,000	,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
KStepMarkov	Pearson Correlation	,999**	,996**	,997**	,948**	,982**	,996**	,996**	,996**	1	,666**	,662**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,000	,000
	N	30	30	30	30	30	30	30	30	30	30	30
MEAN_ subjects	Pearson Correlation	,677**	,670**	,666**	,610**	,682**	,670**	,670**	,670**	,666**	1	,967**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000		,000
	N	30	30	30	30	30	30	30	30	30	30	30
MEDIAN_ subjects	Pearson Correlation	,671**	,670**	,666**	,604**	,672**	,670**	,670**	,669**	,662**	,967**	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	
	N	30	30	30	30	30	30	30	30	30	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Bibliography

[AB02] Adriano Bessa Albuquerque and Arnaldo Dias Belchior. E-commerce website quality evaluation. In *EUROMICRO*, pages 294–300. IEEE Computer Society, 2002.

[aCMIPG] Marc alexis Côté M. Ing, Witold Suryn Phd, and Elli Georgiadou. Software quality model requirements for software quality engineering.

[BCR94] Victor R. Basili, Gianluigi Caldiera, and H. Dieter Rombach. The goal question metric approach. In *Encyclopedia of Software Engineering*. Wiley, 1994.

[Ber01] Michael K. Bergman. The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing*, 7(1), August 2001.

[Bev99] Nigel Bevan. Quality in use: Meeting user needs for quality. *Journal of Systems and Software*, 49(1):89–96, 1999.

[BMB10] Romain Boulet, Pierre Mazzega, and Danièle Bourcier. Network analysis of the french environmental code. In *Proceedings of the 2009 international conference on AI approaches to the complexity of legal systems: complex systems, the semantic web, ontologies, argumentation, and dialogue*, AICOL-I/IVR-XXIV'09, pages 39–53, Berlin, Heidelberg, 2010. Springer-Verlag.

[BMBB02] Nadir Belkhiter, Ghazwa Malak, Mourad Badri, and Linda Badri. Evaluation de la qualité des applications web: état de l'art. In *INFORSID*, pages 107–122, 2002.

[BPCGER08] Bonson-Ponte, Cortijo-Gallego, and Escobar-Rodriguez. Web quality in lithuanian financial institution: a comparaison with the eu results. *ISSN 1392-1258 Ekonomika*, 2008.

[Bra96] Tim Bray. Measuring the web. *Computer Networks*, 28(7-11):993–1005, 1996.

[Bra01] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25:163–177, 2001.

[Bro90] Peter J. Brown. Assessing the quality of hypertext documents. In *ECHT*, pages 1–12, 1990.

[BRS92] Rodrigo A. Botafogo, Ehud Rivlin, and Ben Shneiderman. Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Trans. Inf. Syst.*, 10(2):142–180, 1992.

[BV00] Stuart J. Barnes and Richard T. Vidgen. Webqual: An exploration of web-site quality. In *ECIS*, 2000.

[BW00] Boldyreff and Warren. Websem project: Establishing effective web site evaluation metrics. *WSE'2000*, 2000.

[CCdSP07] Angelica Caro, Coral Calero, Juan Enriquez de Salamanca, and Mario Piattini. Refinement of a tool to assess the data quality in web portals. In *QSIC*, pages 238–243. IEEE Computer Society, 2007.

[CMG+07] Cristina Cachero Castro, Santiago Meliá, Marcela Genero, Geert Poels, and Coral Calero. Towards improving the navigability of web applications: a model-driven approach. *EJIS*, 16(4):420–447, 2007.

[CMM02] Valter Crescenzi, Giansalvatore Mecca, and Paolo Merialdo. Roadrunner: automatic data extraction from data-intensive web sites. In *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, SIGMOD '02, pages 624–624, New York, NY, USA, 2002. ACM.

[Cow00] Adrian Cowderoy. Measures of size and complexity for web-site content, 2000.

[CRS+03] H. Chi, Adam Rosien, Gesara Supattanasiri, A Williams, Christiaan Royer, Celia Chow, Erica Robles, Brinda Dalal, Julie Chen, and Steve Cousins. The bloodhound project: Automating discovery of web usability issues using the infoscent simulator. In *CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters*, pages 505–512. ACM Press, 2003.

[DGM04] Michelangelo Diligenti, Marco Gori, and Marco Maggini. A unified probabilistic framework for web page scoring systems. *IEEE Trans. Knowl. Data Eng.*, 16(1):4–16, 2004.

[FCH+06] X Fang, M. Chau, P. J. Hu, Z. Yang, and O. R. L. Sheng. Web mining-based objective metrics for measuring website navigability. In *International Conference on Information Systems*, 2006.

[FH08] Erik Frøkjær and Kasper Hornbæk. Metaphors of human thinking for usability inspection and design. *ACM Trans. Comput.-Hum. Interact.*, 14(4), 2008.

[FN00] Norman E. Fenton and Martin Neil. Software metrics: roadmap. In *ICSE - Future of SE Track*, pages 357–370, 2000.

[GK06] Marc Guillemot and Dierk König. Web testing made easy. In Peri L. Tarr and William R. Cook, editors, *OOPSLA Companion*, pages 692–693. ACM, 2006.

[GMP95] Franca Garzotto, Luca Mainetti, and Paolo Paolini. Hypermedia design, analysis, and evaluation issues. *Commun. ACM*, 38(8):74–86, 1995.

[HMS+08] May Haydar, Ghazwa Malak, Houari A. Sahraoui, Alexandre Petrenko, and Sergiy Boroday. Anomaly detection and quality evaluation of web applications. *Information Science Publishing; Handbook of Research on Web Information Systems Quality*, 2008.

[HTC95] A. E. Hatzimanikatis, C. T. Tsalidis, and D. Christodoulakis. Measuring the readability and maintainability of hyperdocuments. *Journal of Software Maintenance*, 7:77–90, March 1995.

[Hua03] M. H. Huang. Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, 19(4):425–442, 2003.

[Huy02] P. E. M. Huygen. Use of bayesian belief networks in legal reasoning. presented at the 17 th bileta annual conference, amsterdam 2002. available at http://www.bileta.ac.uk/02papers/huygen.html kadane. In *In 17th BILETA Annual Conference*. John Wiley & Sons, Inc, 2002.

[IEE03] IEEE. Ieee recommended practice for the internet - web site engineering, web site management, and web site life cycle. *IEEE Std 2001-2002 (Revision of IEEE Std 2001-1999)*, pages 01 –72, 2003.

[IH01] Melody Y. Ivory and Marti A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.*, 33(4):470–516, 2001.

[IH02] Melody Y. Ivory and Marti A. Hearst. Improving web site design. *IEEE Internet Computing*, 6(2):56–63, 2002.

[Kal07] James Kalbach. *Designing Web Navigation: Optimizing the User Experience.* O'Reilly Media, 1 edition, 2007.

[KC98] Jurek Kirakowski and Bozena Cierlik. Measuring the usability of web sites. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 42(4):424–428, 1998.

[Ked08] Colombiano Kedowide. Evaluation multicriteres du site web du ministere des ressources humaines et developpement social du canada. Master's thesis, Université de Québec a Montréal, 2008.

[Kee98] Benjamin Keevil. Measuring the usability index of your web site. In *SIGDOC*, pages 271–277, 1998.

[KKTZ05] Muneo Kitajima, Noriyuki Kariya, Hideaki Takagi, and Yongbing Zhang. Evaluation of website usability using markov chains and latent semantic analysis. *IEICE Transactions*, 88-B(4):1467–1475, 2005.

[Kle99] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.

[KSR02] Laurie Kantner, Roberta Shroyer, and Stephanie Rosenbaum. Structured heuristic evaluation of online documentation, 2002.

[KTA06] Christos Katsanos, Nikolaos K. Tselios, and Nikolaos M. Avouris. Infoscent evaluator: a semi-automated tool to evaluate semantic appropriateness of hyperlinks in a web site. In Toni Robertson, editor, *OZCHI*, volume 206 of *ACM International Conference Proceeding Series*, pages 373–376. ACM, 2006.

[KTA10] Christos Katsanos, Nikolaos K. Tselios, and Nikolaos M. Avouris. Evaluating website navigability: validation of a tool-based approach through two eye-tracking user studies. *The New Review of Hypermedia and Multimedia*, 16(1&2):195–214, 2010.

[Ld01] Michael Luck and Mark d'Inverno. A conceptual framework for agent definition and development. *THE COMPUTER JOURNAL*, 44:2001, 2001.

[LH99] David Lowe and Wendy Hall. *Hypermedia & the Web : an engineering approach.* John Wiley & Sons Ltd, Chichester, England, 1999.

[LK06] Younghwa Lee and Kenneth A. Kozar. Investigating the effect of website quality on e-business success : An analytic hierarchy process (AHP) approach. *Decision Support Systems*, 42:1383–1401, 2006.

[LMM04] Pier Luca Lanzi, Maristella Matera, and Andrea Maurino. A framework for exploiting conceptual modeling in the evaluation of web application quality. In Nora Koch, Piero Fraternali, and Martin Wirsing, editors, *ICWE*, volume 3140 of *Lecture Notes in Computer Science*, pages 50–54. Springer, 2004.

[LRA+03] Juan Manuel Cueva Lovelle, Bernardo Martín González Rodríguez, Luis Joyanes Aguilar, José Emilio Labra Gayo, and María del Puerto Paule Ruíz, editors. *Web Engineering, International Conference, ICWE 2003, Oviedo, Spain, July 14-18, 2003, Proceedings*, volume 2722 of *Lecture Notes in Computer Science*. Springer, 2003.

[LS98] Gerald L. Lohse and Peter Spiller. Electronic shopping. *Commun. ACM*, 41(7):81–88, 1998.

[Lyn10] Jonathan Lynn. Internet users to exceed 2 billion this year. *Reuters*, 2010.

[MA07] Stephanos Mavromoustakos and Andreas S. Andreou. Waqe: a web application quality evaluation model. *Int. J. Web Eng. Technol.*, 3(1):96–120, 2007.

[Mal07] Ghazwa Malak. *Evaluation de la Qualité des Applications Web: Approche Probabiliste*. PhD thesis, Université de Montréal, 2007.

[MD01] San Murugesan and Yogesh Deshpande, editors. *Web Engineering, Software Engineering and Web Application Development*, volume 2016 of *Lecture Notes in Computer Science*. Springer, 2001.

[MDHG01] San Murugesan, Yogesh Deshpande, Steve Hansen, and Athula Ginige. Web engineering: A new discipline for development of web-based systems. In Murugesan and Deshpande [MD01], pages 3–13.

[Meh93] Brad Mehlenbacher. Software usability: choosing appropriate methods for evaluating online systems and documentation. In Paul Beam, editor, *SIGDOC*, pages 209–222. ACM, 1993.

[MFC03] Luisa Mich, Mariangela Franch, and G. Cilione. The 2qcv3q quality model for the analysiy of web site requirements. *J. Web Eng.*, 2(1-2):115–127, 2003.

[MFG03] Luisa Mich, Mariangela Franch, and Loris Gaio. Evaluating and designing the quality of web sites. *IEEE MultiMedia*, 10(1):34–43, 2003.

[ML04] Maria De Marsico and Stefano Levialdi. Evaluating web sites: exploiting user's expectations. *Int. J. Hum.-Comput. Stud.*, 60(3):381–416, 2004.

[MRTC06] M. Matera, F. Rizzo, and G. Toffetti Carughi. Web usability: Principles and evaluation methods. In E. Mendes and Mosley N., editors, *Web Engineering*, pages 143–180. Springer Berlin / Heidelberg, 2006.

[MSBB10] Ghazwa Malak, Houari A. Sahraoui, Linda Badri, and Mourad Badri. Modeling web quality using a probabilistic approach: An empirical validation. *TWEB*, 4(3), 2010.

[MVBM04] Céline Mariage, Jean Vanderdonckt, Abdo Beirekdar, and Noirhom Monique. Destine: outil d'aide a l evaluation de l ergonomie des sites web. In *Proceedings of the 16th conference on Association Francophone d'Interaction Homme-Machine*, IHM 2004, pages 117–124, New York, NY, USA, 2004. ACM.

[Nie95] Jakob Nielsen. Usability inspection methods. In *CHI 95 Conference Companion*, pages 377–378, 1995.

[NL06a] J. Nielsen and H. Loranger. *Prioritizing Web usability*. Voices Series. New Riders, 2006.

[NL06b] Jakob Nielsen and Hoa Loranger. *Prioritizing Web Usability*. New Riders, 2006.

[NM90] Jakob Nielsen and Rolf Molich. Heuristic evaluation of user interfaces. In Jane Carrasco Chew and John A. Whiteside, editors, *CHI*, pages 249–256. ACM, 1990.

[NMVN05] Shivaram Narayanan, Madhav Marathe, Anil Vullikanti, and Shivaram Narayanan. A study of betweenness centrality on biological networks, 2005.

[NWL$^+$04] P. Naïm, P-H. Wuillemin, P. Leray, O. Pourret, and A. Becker. *Réseaux bayésiens*. Eyrolles, Paris, 2004.

[Off02] A. Jefferson Offutt. Quality attributes of web software applications. *IEEE Software*, 19(2):25–32, 2002.

[OLR01] Luis Olsina, Guillermo Lafuente, and Gustavo Rossi. Specifying quality characteristics and attributes for websites. In Murugesan and Deshpande [MD01], pages 266–278.

[OPL$^+$03] María E. Alva O., Ana Belén Martínez Prieto, Juan Manuel Cueva Lovelle, T. Hernán Sagástegui Ch., and Benjamín López. Comparison of methods and existing tools for the measurement of usability in the web. In Lovelle et al. [LRA$^+$03], pages 386–389.

[Pal02] Jonathan W. Palmer. Web site usability, design, and performance metrics. *Information Systems Research*, 13(2):151–167, 2002.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web, 1999.

[PC95] Peter Pirolli and Stuart Card. Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '95, pages 51–58, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[PF03] Peter Pirolli and Wai-Tat Fu. Snif-act: A model of information foraging on the world wide web. In Peter Brusilovsky, Albert T. Corbett, and Fiorella de Rosis, editors, *User Modeling*, volume 2702 of *Lecture Notes in Computer Science*, pages 45–54. Springer, 2003.

[Pre05] Roger S. Pressman. *Software Engineering, a Practitioner's Approach*. McGraw-Hill, 6th edition, 2005.

[RCP03] Julián Ruiz, Coral Calero, and Mario Piattini. A three dimensional web quality model. In Lovelle et al. [LRA$^+$03], pages 384–385.

[RT06] Filippo Ricca and Paolo Tonella. Detecting anomaly and failure in web applications. *IEEE MultiMedia*, 13:44–51, April 2006.

[SD02] Petra Schubert and Walter Dettling. Extended web assessment method (ewam) - evaluation of e-commerce applications from the customer's viewpoint. In *HICSS*, page 175, 2002.

[SDKP06] Ahmed Seffah, Mohammad Donyaee, Rex Bryan Kline, and Harkirat Kaur Padda. Usability measurement and metrics: A consolidated model. *Software Quality Journal*, 14(2):159–178, 2006.

[SGM00] Houari A. Sahraoui, Robert Godin, and Thierry Miceli. Can metrics help to bridge the gap between the improvement of oo design quality and its automation? In *ICSM*, pages 154–162, 2000.

[SM97] Simon Buckingham Shum and Cliff Mcknight. World wide web usability: introduction to this special issue, 1997.

[Ste01] Darren Stephens. Web engineering literature review - version 0.05. *University of Hull*, December 2001.

[Suh05] Woojong Suh. *Software Engineering, Principles and Techniques.* Idea Group, 2005.

[SVC10] G Sreedhar, Rashtriya Sanskrit Vidyapeetha, and National Informatics Centre. Measuring quality of web site navigation 1 1. *Science*, pages 80–86, 2010.

[SX08] Antonia Stefani and Michalis Nik Xenos. E-commerce system quality assessment using a model based on iso 9126 and belief networks. *Software Quality Journal*, 16(1):107–129, 2008.

[TSPN08] D. Tedesco, A. Schade, K. Pernice, and J. Nielsen. Site map usability. technical report. *Nielsen Norman Group*, 2008.

[VBSH09] Stéphane Vaucher, Samuel Boclinville, Houari A. Sahraoui, and Naji Habra. Recommending improvements to web applications using quality-driven heuristic search. In Gottfried Vossen, Darrell D. E. Long, and Jeffrey Xu Yu, editors, *WISE*, volume 5802 of *Lecture Notes in Computer Science*, pages 321–334. Springer, 2009.

[VHK10] Benoit Vanderose, Naji Habra, and Flora Kamseu. Towards a model-centric quality assessment. *IWSM/MetriKon*, 2010.

[VS10] Stéphane Vaucher and Houari A. Sahraoui. Multi-level evaluation of web site navigability. In *WISE*, 2010.

[Wal06] Steven Walfish. A Review of Statistical Outlier Methods. *Pharmaceutical Technology*, November 2006.

[Whi03] Katy Whitelaw. Why make websites accessible?: and how? In Lynnell Lacy, William S. Thieke, and Gail Farally-Semerad, editors, *SIGUCCS*, pages 259–261. ACM, 2003.

[WL03] Richard Wheeldon and Mark Levene. The best trail algorithm for assisted navigation of web sites. *CoRR*, cs.DS/0306122, 2003.

[WRH+00] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Bjöorn Regnell, and Anders Wesslén. *Experimentation in software engineering: an introduction.* Kluwer Academic Publishers, Norwell, MA, USA, 2000.

[WS03] Scott White and Padhraic Smyth. Algorithms for estimating relative importance in networks. In Lise Getoor, Ted E. Senator, Pedro Domingos, and Christos Faloutsos, editors, *KDD*, pages 266–275. ACM, 2003.

[WW04] Harold W. Webb and Linda A. Webb. Sitequal: an integrated measure of web site quality. *The Journal of Enterprise Information Management*, 17(6):430–440, 2004.

[ZL06] Yi Zhang 0012, Lei Zhang 0013, Yan Zhang, and Xiaoming Li. Xrank: Learning more fromweb user behaviors. In *CIT*, page 36. IEEE Computer Society, 2006.

[ZLW07] Yuming Zhou, Hareton Leung, and Pinata Winoto. Mnav: A markov model-based web site navigability measure. *IEEE Trans. Software Eng.*, 33(12):869–890, 2007.

[ZS06] Sven Ziemer and Tor Stålhane. Web application development and quality - observations from interviews with companies in norway. In José A. Moinhos Cordeiro, Vitor Pedrosa, Bruno Encarnação, and Joaquim Filipe, editors, *WEBIST (1)*, pages 495–498. INSTICC Press, 2006.

[ZZ07] Hong Zhu and Yanlong Zhang. Navigability design and measurement. In Arthur Tatnall, editor, *Encyclopedia of Portal Technologies and Applications*, pages 642–646. IGI Global, 2007.

[ZZG04] Yanlong Zhang, Hong Zhu, and Sue Greenwood. Website complexity metrics for measuring navigability. In *QSIC*, pages 172–179. IEEE Computer Society, 2004.